



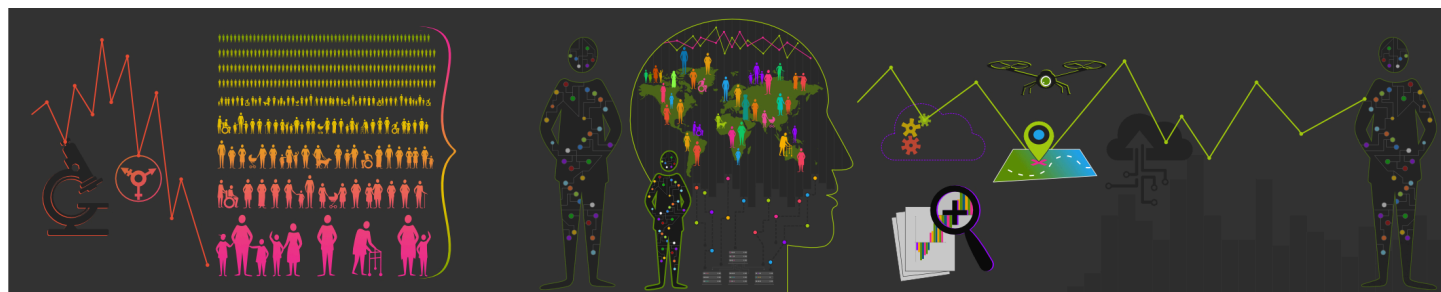
Living with Data survey report

June 2021

Helen Kennedy, Mark Taylor, Susan Oman, Jo Bates, Itzelle Medina-Perea, Hannah Ditchfield, Lulu Pinney, 16/06/2021

Contents

1. Introduction
2. About the survey
3. Summary of findings
4. About respondents
5. Awareness of data uses
6. General attitudes to data uses
7. Comparing concerns about data uses with other concerns
8. Trust in institutions' data uses
9. Attitudes to specific public sector data uses
10. Classifying awareness and attitudes
11. Conclusions & recommendations
12. References
13. Appendices.



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. Visit www.nuffieldfoundation.org.

1. Introduction

Living With Data is a research project funded by the Nuffield Foundation, 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 capturing these processes. The data at the centre of such practices is often personal data, which has been 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 rights¹).

This document reports on a survey of perceptions of data uses, undertaken with 2000 adults in the UK in September and October 2020, via the web survey platform Qualtrics. This report describes what we found. Interpretations of our findings and recommendations for policy, practice and further research can be found in a summary of the survey findings which can be accessed at <https://livingwithdata.org/current-research/publications/>. Qualitative focus group and interview research we are also undertaking is reported elsewhere.

Increasingly widespread data uses result in harms as well as benefits. Concern about potential harms has led to initiatives which aim to influence data governance and usage, so that data works 'for people and society' (the mission of the Ada Lovelace Institute) and is 'a force for good' (an aim of the UK government's Centre for Data Ethics and Innovation (CDEI)). This concern also motivates our research, as we believe that understanding how people perceive data uses can inform efforts to minimise harm. There has been a growth in research into public perceptions of data practices in recent years, but there are some gaps in research and understanding, which we aim to fill. These are:

- A lot of research into public perceptions of data practices has focused on commercial organisations. **Public sector data practices** increasingly shape everyday life experiences, so we also need to research perceptions of public sector data uses. This has started to happen, but more research is needed on **differences** across sectors.
- Much existing research focuses on attitudes to high-profile or contentious data practices. We also need to understand people's views of **everyday data practices**, such as those that take place in the public sector, because of the role that they play in shaping everyday lives.
- In survey-based research, there is **limited analysis**. Often, descriptive statistics about whole samples are presented, with inequalities and demographic differences only occasionally addressed. More detailed analysis is needed of whether and how perceptions and attitudes vary across different groups. This is important because the harmful consequences of data uses are not the same for all groups.
- More detailed analysis is needed in order to understand the **relationship between awareness or understanding of data uses and attitudes towards them**. For example, are people who are more knowledgeable more or less concerned about data uses? This matters when it comes to building trust in institutions' uses of data.
- Most research explores attitudes towards data uses in general or in the abstract. We believe that in order to improve data uses, we need to understand people's perceptions of **real-world, specific data practices in specific domains**. Researching this will help us understand the role that context plays in perceptions of data uses, which has been identified as important, but which is not widely researched (Kennedy et al 2020).

¹ European Union (2016) *General data protection regulation*, Off J Eur Union 49: L119. <https://gdpr-info.eu>.

- Research has found that **social inequalities** play a major role in shaping people's *experiences* of data uses. There is little research into the relationship between social inequalities and *perceptions* of data uses, especially quantitative research. Research which helps us understand the relationship between social inequalities and perceptions of data uses will help us identify whether and how data uses can be improved.
- Understanding what kinds of data uses people consider to be **fair or ethical** is also important, in order for debate about 'good' data uses to be informed by public views.

2. About the survey

2.1 Sample

Data was collected from the 18th of September 2020 to the 15th of October 2020, via the web survey platform Qualtrics. Participants were recruited by Qualtrics. The sample was recruited to be nationally representative of adults in the UK, in relation to gender, age, income and ethnicity. Numbers of disabled respondents were also nationally representative. There was additional recruitment (or 'boosts') of people born outside the UK, LGBTQ+ people, and people in receipt of Universal Credit, to ensure these groups were large enough for analysis and that the views of people in these groups were represented in our study. We did not oversample in relation to educational qualifications because this landscape is complex; as a result, our respondents have slightly higher qualifications than is nationally representative. Respondents whose answers suggested they were not paying attention were filtered from the eventual sample both during and after the data collection process, leaving an overall sample size of 2,000.

2.2 Survey design

The survey consisted of seven main parts.

1. **Personal characteristics:** Respondents answered questions about their personal characteristics, such as their gender, employment status, and disability. These questions are important to understand the role of inequalities in perceptions of data uses. The data collected also enabled the survey platform to manage recruitment, and more general eligibility for the study. For example, participants who responded that they lived outside of the UK, or who responded that they were 17 or younger, were screened out.
2. **Internet usage:** Respondents answered questions about their current use of the internet, the frequency with which they use the internet, with what devices, what they do online, and how confident they felt online. We asked these questions in order to analyse whether these factors influence attitudes to data uses.
3. **Awareness of data uses:** Respondents were asked to give yes/no/don't know or true/false/don't know answers to questions about data uses, to enable us to analyse whether awareness, knowledge and understanding of how data is used influence attitudes to data uses.
4. **General attitudes to data uses:** Respondents answered questions about their attitudes towards general data uses, to enable us to analyse whether these attitudes influence attitudes to the specific data uses we presented later in the survey.
5. **Trust in institutions' data uses:** Respondents answered questions about their trust in various different institutions to a) keep their data safe, b) gather and analyse data about them in responsible ways, and c) be open and transparent about what they do with data. We asked these questions to gauge whether trust varies across institutions and data practices, and whether trust in BBC, DWP and NHS to do these things influences attitudes to the specific BBC, DWP and NHS data uses we presented later in the survey.

6. **Attitudes to specific public sector data uses:** We asked about attitudes to specific data uses that we described to them. Respondents were randomly allocated to one of three batches of questions, about current or potential future BBC, DWP or NHS data practices. So 1/3 of the 2000 respondents were asked to respond to questions about the specific data practices of each organisation. Questions for each data practice were different because the practices themselves differed, but each batch concluded with questions about how much people felt they understood the data practices, and how surprised they were to read about them. Respondents were also given the option of providing more detail in a free text field about why they would/would not use a data-driven service or why they were/were not comfortable with a data use.
7. **Concerns about data uses as they compare with other concerns:** We asked respondents about what, if anything, concerns them in their daily lives, in order to understand the extent of any concerns expressed about data practices in the survey. We asked how concerned people were about two given lists of items: a list of 15 general concerns (eg the economy, healthcare, derived from Ipsos Mori's Most Important Issue tracker) and 13 specific concerns (eg funding for the NHS, anti-social behaviour and crime in local area, data being used in unfair ways).

3. Summary of findings

3.1. Awareness of data uses

- Respondents demonstrated awareness of some data uses, with most people correctly answering a number of simple questions about how data are collected and used. When asked more detailed questions, such as how different data sources are combined, or what it means to have a privacy policy, a majority of respondents either answered incorrectly or stated that they didn't know the answer.
- The group most likely to report that they were not sure about whether the different factual statements about data uses that we presented to them were true or false was those with low qualification levels. This group was also more likely to classify statements incorrectly.
- Women were more likely to state that they didn't know whether factual statements were true or false, but no less likely than men to classify them correctly.
- It is important to acknowledge that it can be difficult to know about, understand or be aware of data uses, because they are not transparent, or they change frequently. We should therefore be careful about how we interpret limited awareness or understanding of data uses.

3.2. Attitudes towards and concerns about data uses

- People are concerned about data uses. Respondents indicate high levels of concern about data uses in responses to questions which invited them to say how much they agreed with attitudinal statements. High levels of concern have been identified in previous research, including our own. This consistent finding communicates a strong message to data policy-makers and practitioners.
- Respondents want to know who has access to data about them, they want more control over how their data is used by organisations, and they want to know where data about them is stored. They do not support corporate profit-making from personal data, and only a minority of respondents don't have strong opinions about the collection and use of data about them.
- Context is a defining factor when it comes to attitudes to data uses. Some data uses are more concerning than others and there are important differences in attitudes, depending on the type of data use. It is not helpful to ask people about their attitudes to data uses out of context.

- In general, attitudes towards data uses were largely similar across groups, with some small differences. Disabled people were more positive about the re-use and sharing of health data for research purposes, and more concerned by commercial companies providing data-driven public services than other groups. Women wanted knowledge and control of their personal data more than men. People with university qualifications were more likely to support the re-use and sharing of health data for research, and to believe that collecting and analysing data can be good for society, than people without university qualifications.
- Respondents are fairly or very concerned about data practices. Compared to other *narrow* concerns, only the economic costs of Covid-19 and funding for the NHS ranked more highly than data being used in unfair ways. Commercial companies profiting from personal data, organisations tracking when, where and how people log on, and automated technologies being used to try to change people's behaviours were also ranked highly.
- Personal data being used to manage Covid-19 was the least concerning from a list of specific concerns. This shows the importance of context in concerns about data uses.
- Younger people, people of colour, LGBTQ+ people, and people in receipt of Universal Credit were more concerned about narrow data uses than other groups, though differences were never very large.

3.3. Trust in organisational data uses

- Respondents trust public health care professionals most with their data, and media and social media companies least. High levels of trust in the healthcare sector and low levels of trust in the BBC were noteworthy. The most significant differences across groups relate to age: older people had more trust in their GP, the police, and the DWP, and less trust in social media companies.
- Levels of trust expressed by respondents are consistent across the three data uses we asked about: keeping data safe, gathering and analysing data in responsible ways, and being open and transparent about what is done with data. This consistency suggests that respondents' views of and trust in sectors and institutions strongly influences their trust in the same sectors' and institutions' data uses. In other words, context matters.

3.4. Attitudes towards specific public sector data practices

We asked questions about attitudes to a) two BBC experiments with personal control over data uses, b) the NHS Covid data store, and c) a DWP identity verification process for Universal Credit payments. There were differences in understanding of, degree of surprise about, likelihood to use and comfort with the data uses that we presented, by income, education, and age, with higher-paid, higher-qualified, and younger people reporting less surprise and more understanding.

- Respondents were unlikely to use the **BBC experiment** that was targeted at adults (BBC Box), especially older people, people with lower qualifications, white people, and men. More respondents (60%) reported feeling comfortable with a child they know using the app targeted at children (Own It). They were surprised to find out about BBC experiments.
- Around three out of four respondents who answered questions about the **NHS Covid data store** said they felt comfortable with it. 58% of respondents said they were not surprised to learn details about it, and 34% were surprised. The vast majority (70%) of respondents understand the datastore somewhat. LGBTQ+ respondents expressed more discomfort with the NHS Covid data store than heterosexual cisgender respondents.
- Respondents were largely comfortable with the **DWP's use of identity verification in relation to Universal Credit**. They were less comfortable with the possibility of DWP checking whether people logging into accounts swipe their phones in the same pattern as they normally swipe it as a future

extra security check. People in receipt of Universal Credit were more comfortable with the checks described than people not in receipt of Universal Credit. Higher-paid people were more likely to say they would use the identity verification option if claiming Universal Credit.

3.5. Types of awareness and attitudes and how these relate to each other and to other factors

In our analysis, we grouped people according to their responses to awareness and attitude questions, in order to understand how these relate to each other and to other factors. We produced the classifications listed below based on our analysis of responses, not on respondents' self-descriptions.

- We find four clusters of awareness, which we call:
 - *Knowledgeables*: respondents who generally correctly identify true and false statements;
 - *Believers*: generally respond that the statements they are presented with are true;
 - *Disbelievers*: generally respond that statements are false, even when they are in fact true;
 - *Don't knows*: generally state that they don't know the answer to awareness questions.
- We find four clusters of attitudes, which we call:
 - *Critical*: respondents whose responses are strong, strongly disagreeing with some statements and strongly agreeing with others;
 - *Cautious*: tend to agree or disagree but do not respond strongly;
 - *Neutral*: overwhelmingly respond with 'Neither agree nor disagree';
 - *Agree*: overwhelmingly either agree or strongly agree with statements.
- *Knowledgeables* are more likely to have more critical attitudes towards data uses. *Disbelievers* are less likely to have critical attitudes towards them. *Don't knows* are also less likely to hold strong opinions about data uses. *Knowledgeables* were least likely to report that they were surprised by the data uses described. *Disbelievers* were the most likely to be comfortable with data uses, while *Knowledgeables* and *Critical* respondents were less comfortable. *Knowledgeables* are no more likely than other groups to report that they understand the data uses described.
- The largest differences in awareness clusters are related to qualifications, where people with higher qualifications are less likely to be *Disbelievers* and *Don't knows*. The largest differences in attitude clusters are related to age, where older people are much more likely to be *Critical*. Older people were more likely to be *Believers* and less likely to be *Disbelievers*. Younger people were much more likely to be in the *Agree* and *Neutral* clusters, and older people more likely to be in the *Critical* cluster. People with higher qualifications were more likely to be *Knowledgeables*, and much less likely to be *Don't knows*. Black people were more likely to be in the *Cautious* attitude cluster. Men were slightly more likely to be *Knowledgeables* than women and much less likely to be *Don't knows*, and they were also more likely to be in the *Critical* attitude cluster.

4. About respondents

4.1 Respondent characteristics

Tables 1 and 2 summarise our respondents' characteristics, presented in the order we asked the questions. The first questions asked about employment, household sources of income, and their overall annual household income. We asked about household sources of income to ascertain if respondents were in receipt of Universal Credit (UC), as one of the example data uses that we presented later in the survey related to UC and we wanted to assess whether experience of claiming UC influenced responses

to the related data use. The sources of income are not mutually exclusive, and so will not sum to 100%. Other categories may not sum to 100% due to rounding.

The demographic characteristics of respondents are similar to those found in nationally representative face-to-face surveys, except where we boosted numbers, as described above, and in relation to education, as our sample has a larger proportion of people with university qualifications.

In the subsequent analysis, we compare different groups, focusing on age, gender, qualifications, disability, sexuality, ethnic group, and income, where the sample size allows. Because some questions were only asked to a subsample of respondents, we aggregate groups together: this is the case for the questions on specific public sector data uses (discussed in Section 9), where we group together people from Black, Asian, Mixed, and Other ethnic groups into the category ‘people of colour’. While some of our respondents are non-binary, we do not make statements about our non-binary participants as their numbers in the sample are too small to draw inferences.

Table 1: employment and income data

Variable	Category	Proportion
Activity	Working full-time	44%
	Working part-time	15%
	Unemployed and looking for work	6%
	Furloughed	2%
	Full-time university student	3%
	Other full-time student	1%
	Retired	16%
	Not in paid work for any other reason	9%
	Other	3%
Source of income	Earnings, wages, salary, bonuses	67%
	Income from self-employment	12%
	Occupational pensions, state retirement pensions	20%
	Universal Credit	12%
	Any other state benefits	12%
	Interest from savings or investments	13%
	Rent from property	4%
	Other income	4%
	None of these	3%
Total household income	£19,999 or less	29%
	£20,000 to £39,999	38%
	£40,000 to £59,999	21%
	£60,000 to £100,000	10%
	£100,000 or more	2%

Table 2: demographic data

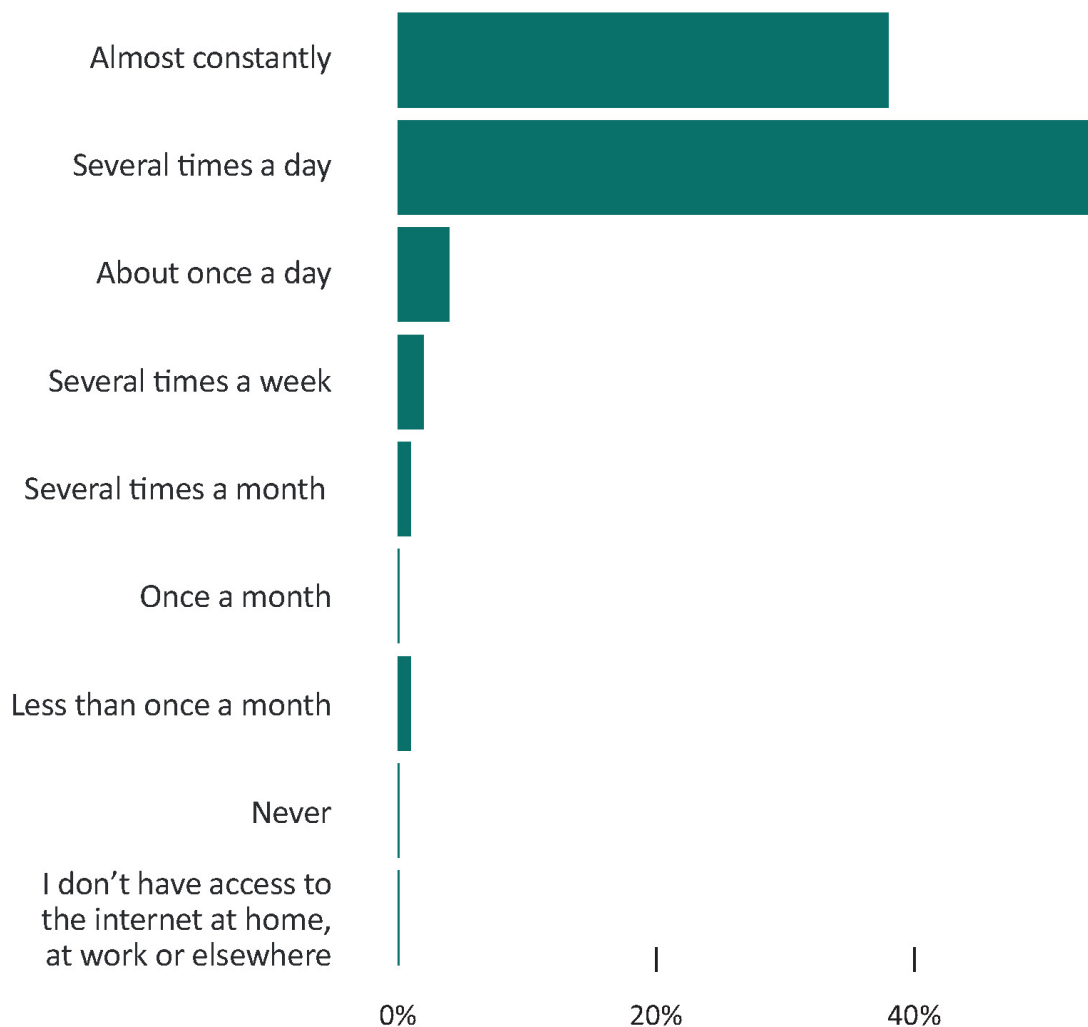
Variable	Category	Proportion
Highest education qualification	None / don't know	6%
	Post-16 qualifications (eg GCSEs)	27%
	Post-18 qualifications (eg A-levels)	19%
	Some higher education (eg nursing qualification)	8%
	Undergraduate degree	23%
	Higher degree	10%
	Other qualification	6%
Age	18-24	12%
	25-34	19%
	35-44	18%
	45-54	20%
	55-64	17%
	65 or older	14%
Adults in household	1	24%
	2	53%
	3 or more	23%
Children 0-5 in household	0	88%
	1	11%
	2 or more	1%
Children 6-10 in household	0	83%
	1	14%
	2 or more	3%
Children 11-17 in household	0	82%
	1	14%
	2 or more	4%
Country of birth	UK	87%
	Other	13%
Citizenship	UK	93%
	Other European country	4%
	Other non-European country	4%
Disability	Reported disability	27%
Ethnic group	White British	72%
	White Other	5%
	Black	6%
	Asian	11%
	Mixed ethnicities / other	6%
Gender	Male	48%
	Female	51%

Sexuality	Non-binary	1%
	Heterosexual / straight	80%
	LGBTQ+	20%

4.2 Respondent internet usage

We asked about internet usage, to explore the relationship between frequency of and confidence in internet usage, awareness of and attitudes towards data uses. The majority of participants (96%) use the internet at least daily, which is to be expected given the recruitment method was via an online survey platform. Although we did recruit a number of participants who use the internet less frequently, most respondents are more regular users of the internet than the general population. Figure 1 shows the frequency with which our participants use the internet.

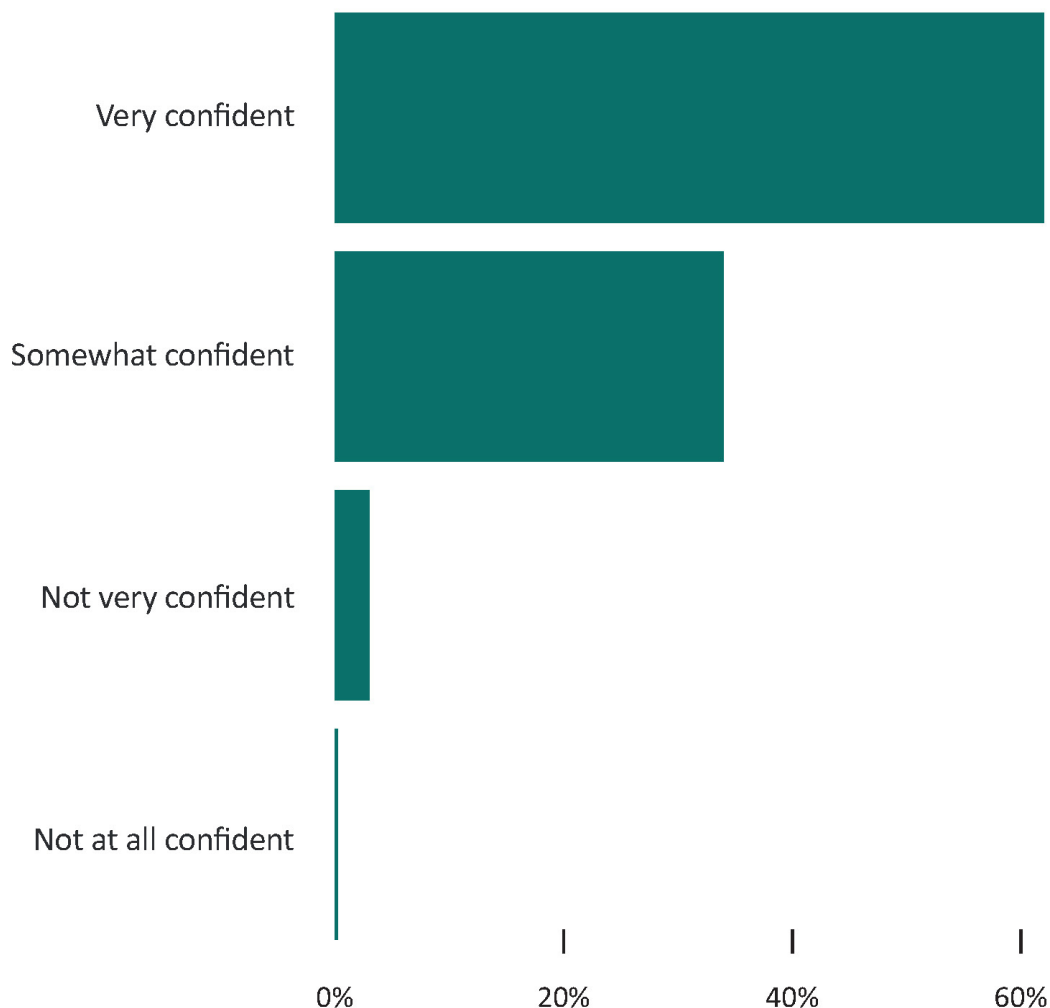
Figure 1: How often do you use the internet for your personal use?



The most common way people access the internet is through a mobile device (87%), followed by through a computer (83%). 45% of respondents access the internet via a TV and 23% via a games console. Just 4% of respondents access the internet in other ways. Respondents selected all responses that applied, so responses do not sum to 100%.

The most common online activity for respondents is sending and receiving emails, at 90%, followed by purchasing goods online (80%) and using social network services (77%). Smaller fractions have uploaded content, signed petitions, or posted opinions on civic or political issues. Finally in this section, we asked people how confident they were doing the things they need to do online. 62% described themselves as very confident and 34% somewhat confident, with 3% not very confident and less than 1% not at all confident. This is illustrated in Figure 2.

Figure 2: Overall, how confident do you feel using computers, smartphones, or other electronic devices to do the things you need to do online?



4.3. Demographic differences in internet use and confidence

The largest differences in respondents' internet use, what they do online and how they do it, relate to age. Younger respondents are more likely to be online almost constantly, and frequency of use decreases with age. The fractions of people using a games console or a TV to access the internet also decreases with age, and the oldest age group is less likely to use a mobile device than other age groups.

There are also some differences in what people of different age groups do online. The proportion using the internet to send and receive emails increases with age, as does the proportion using the internet to find information about goods and services. The fraction uploading content, telephoning, and using social network sites is lower among older groups. Finally, there are age-related differences in online confidence. The younger groups, up to and including the 35-44 group, feel very confident online: the 25-34 group is most confident, with 79% rating themselves as very confident. Smaller fractions of the older groups are very confident online, particularly the group aged 65 and older, where just 38% rate themselves as very confident online.

There are similar differences by education and income, so we only highlight differences amongst people of different qualification levels here. First, there are differences in the devices people use. People with no qualifications or vocational qualifications are less likely to have a mobile device (73% and 76% respectively), while those people with more qualifications are more likely to have a computer (89%). In terms of online activity, the major difference is that people with more qualifications are more likely to use the internet for telephone and video calls; otherwise, the differences are small. People with higher qualifications are more likely to be very confident online (74% of people with degrees, compared with 53% of people with no qualifications). The differences in internet use relating to age, education and income may inform answers that respondents gave to subsequent questions.

There are no major differences between men and women, between people of different ethnic groups, or between cisgender heterosexual people and LGBTQ+ people in their internet use.

5. Awareness of data uses

5.1. What constitutes awareness, and how does it relate to knowledge and understanding?

What does it mean to be aware of data uses? As has been widely established (eg Kitchin 2014), data collection, analysis, sharing and use are often opaque and hard to know about or understand. We derived some of our questions in this section of the survey from other surveys which provided answers to said questions (such as Doteveryone, 2018 and ICO, 2019). Yet in our team, we disputed whether some statements, identified as true or false in those other surveys, were in fact so. For example, we were not all convinced that data is used to protect people from scams, a statement that other surveys categorise as true. Furthermore, understanding of data uses is not static – it can change in the process of thinking or reading about data uses. It is important, therefore, not to oversimplify what it means to know about or understand data uses. We use the term 'awareness of data uses' here, rather than knowledge, as this term captures knowledge that might people have about data uses without full understanding of their precise details².

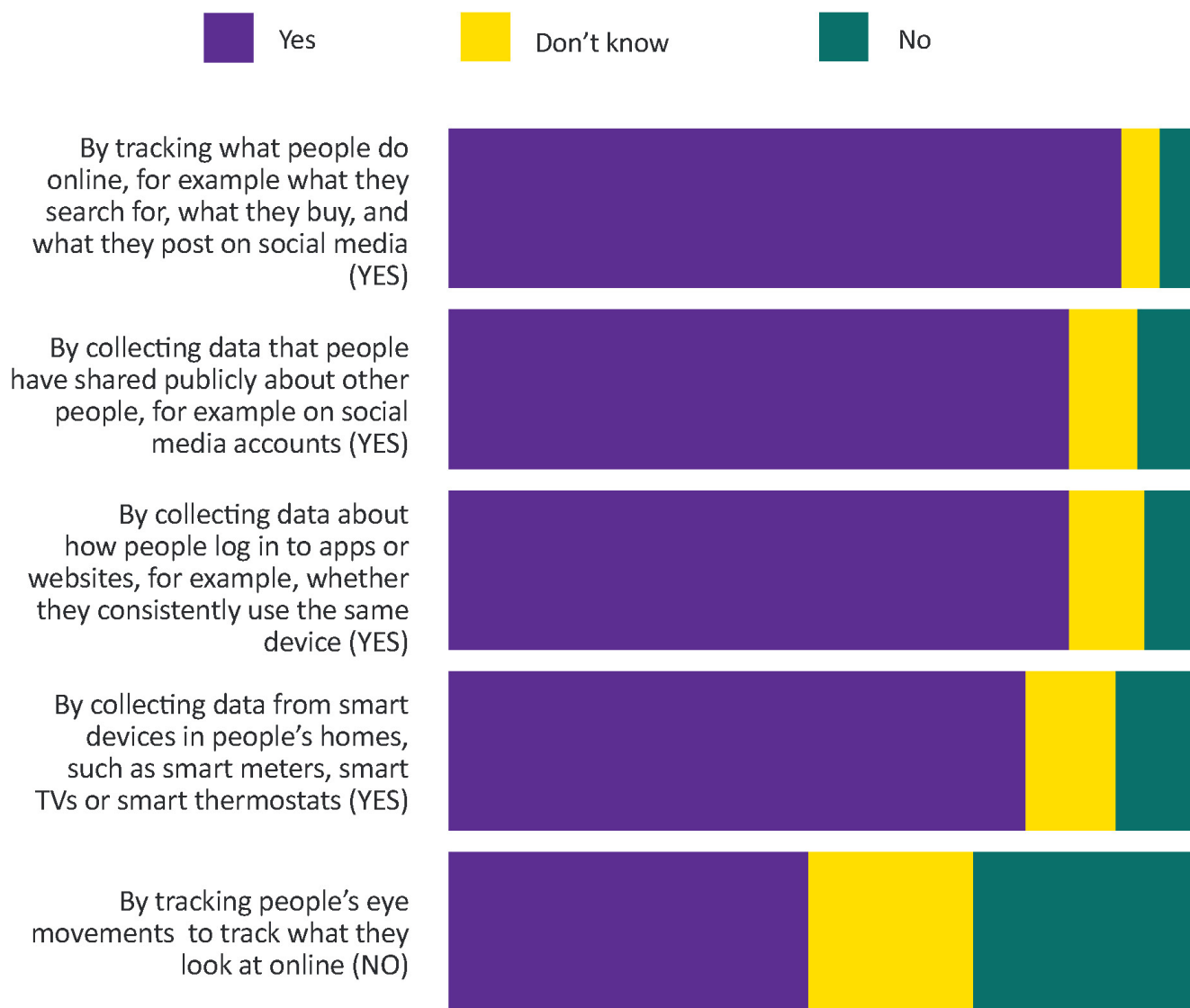
² We thank Aidan Peppin of the Ada Lovelace Institute for this suggestion.

5.2. Overarching findings

We gauged respondents' awareness of data practices, by asking them to provide yes, no or don't know answers to a question about whether they thought that data is *collected* in any of the ways we listed, and then by asking if they thought that data is *used* in ways listed. We then presented respondents with nine statements about data uses and asked them to indicate whether they thought that the statements were true, false, or that they didn't know. The order that statements were presented was randomised.

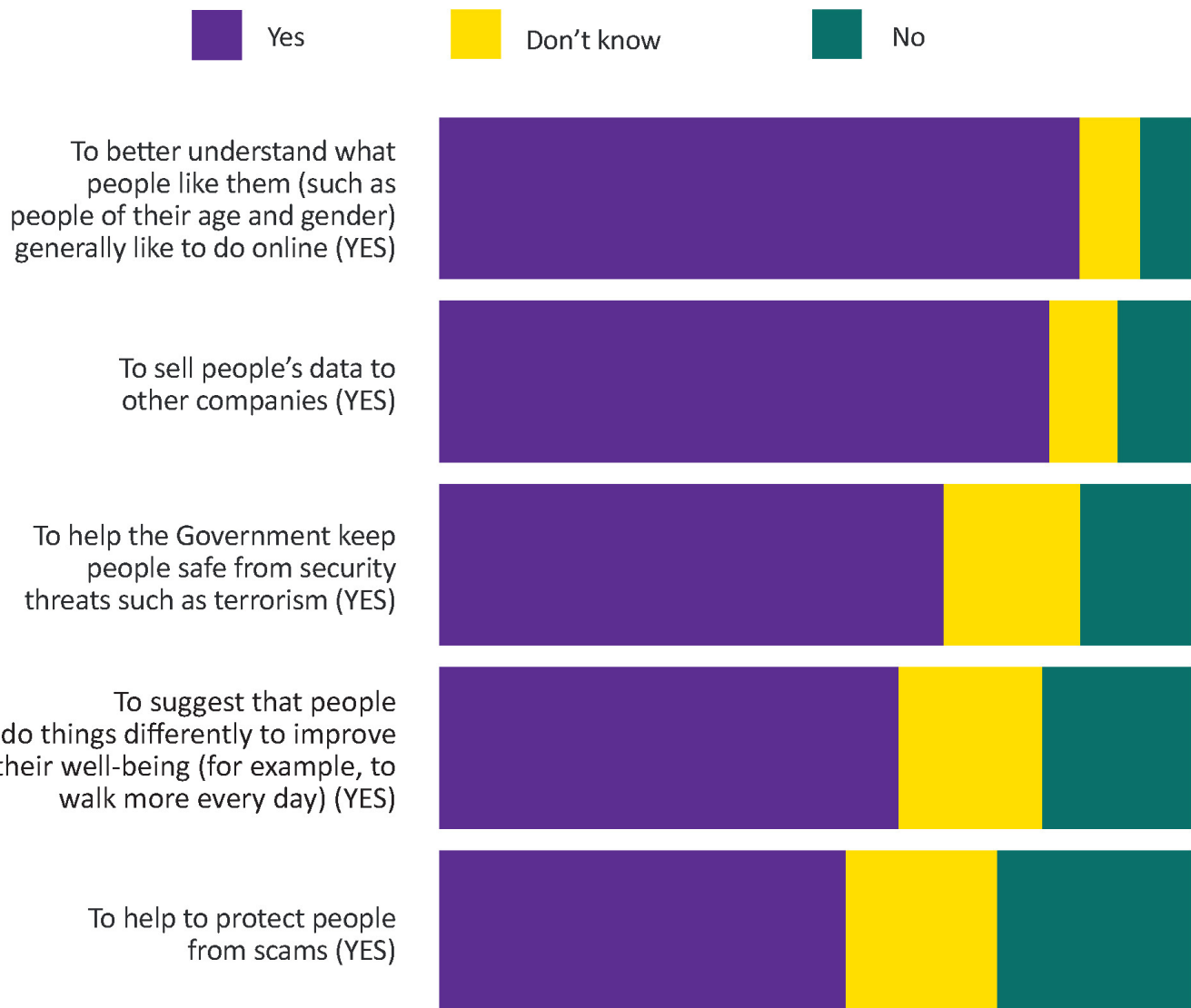
A majority of respondents correctly identified the true statements about data collection as such, while 48% believed the false statement (that organisations collect data by tracking people's eye movements to track what they look at online) was true. In most cases, there were similar numbers of people responding that they didn't know and that statements were false. Results are shown in Figure 5.

Figure 5: Do you think that any organisations **collect** data about people in the following ways?



In the next question about data uses, where all five statements were true, the majority of respondents correctly identified them as such. Larger fractions responded that they thought that the latter three data uses either didn't occur or that they didn't know whether they occur. Results are shown in Figure 6.

Figure 6: Do you think that any organisations *use* data about people in the following ways?



In the true/false questions, three true statements were correctly identified as such by a majority of respondents. These were about: Facebook linking activity across devices; the definition of personal data; and organisations being required by law to reveal the data they hold about individuals. The other true statement, 'It's possible for data about me from different services to be combined and stored on a device in my house' was asked because later in the survey, we explored respondents' views about two BBC experiments with personal data stores (PDS, explained below) which enable this, and we wanted to be able to compare awareness of this phenomenon with attitudes to specific examples of PDS in action. This true statement was identified as such the least frequently, with 39% 'don't know' responses.

Of the five false statements, only one, about banks sending emails with verification links, was correctly identified as such by a majority of respondents. The three other false statements, about apps being legally forbidden to share location data with others, free WiFi providers having to secure the service, and automated decision-making being always less biased than human decision-making, had similar numbers of people responding true, false, and don't know. A majority of respondents incorrectly believed that when a website has a privacy policy, this means it will not share people's data with other websites or companies without their permission. Results are shown in Figure 7.

Figure 7: Are the following statements true or false?



Respondents demonstrated some awareness of some data uses. Respondents were least likely to correctly answer questions about particular data uses, such as how different data sources are combined, or what it means to have a privacy policy. Respondents were more likely to state that false statements were true than vice versa: 3/4 of true statements were correctly identified as such by a majority of respondents, compared to only 1/5 of false statements. Respondents were more likely to respond that they didn't know whether false statements were true or false, indicating a reluctance to state categorically that data uses *do not* occur. This in turn suggests that respondents believe that a broad range of data practices *may* happen, even if they do not know for certain that particular practices do. Thus we conclude that there is broad awareness of data uses in general.

5.3. Demographic differences in awareness of data uses

There were few differences in responses to awareness questions across demographic groups. Differences existed in the fractions of people responding 'don't know' rather than in those classifying false statements as true, or vice versa. The group most likely to report uncertainty was respondents with lower levels of qualifications. For example, 16% of people with no qualifications responded that they weren't sure if any organisations sold people's personal data to other companies, compared with 7% of people with degrees. This group was also more likely to classify statements incorrectly. Women were more likely to state that they did not know the answers to awareness questions than men, but no less likely than men to correctly answer questions. Women gave more 'don't know' answers, men gave more incorrect answers, suggesting a gender difference in relation to confidence rather than knowledge or awareness which has been identified in research in a wide range of fields.

6. General attitudes to data uses

6.1. Overarching findings

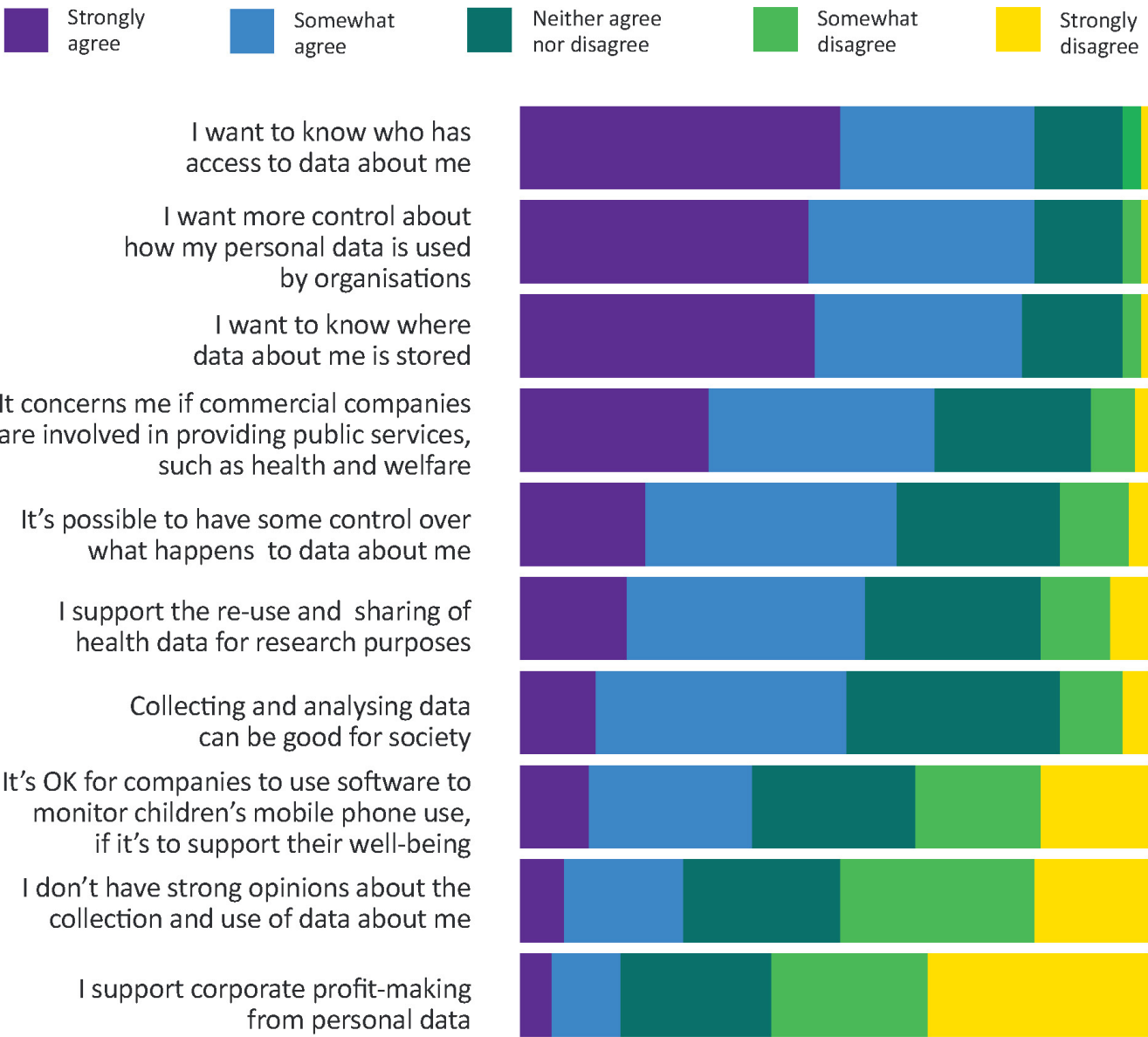
To gauge respondents' attitudes to data practices, we presented them with ten statements and asked them to indicate how much they agreed or disagreed with each one. These statements were prefaced with a note that encouraged respondents to answer honestly and noted that there are no right or wrong answers to these questions. As in the section of the survey discussed in (5) above, we included some statements that were linked to the specific public sector data practices that we examined later, so that we could compare attitudes to these general statements with attitudes to specific data practices. These related to re-using health data for research purposes, the involvement of commercial companies in the provision of public services, and monitoring children's mobile phone use to support their well-being. All statements and responses are presented below in Figure 8.

Respondents indicated high levels of concern about data uses. High levels of concern have been identified in previous research, including our own (Kennedy et al 2020). For example, respondents want to know who has access to data about them (83% agree/strongly agree with the relevant statement), they want more control over how their data is used by organisations (83%), and they want to know where data about them is stored (80%). They do not support corporate profit-making from personal data (60% disagree/strongly disagree with the relevant statement), and only 26% of respondents don't have 'strong opinions about the collection and use of data about me'. This consistent finding communicates a

strong message to data policy-makers and practitioners: people are not happy with current data practices and infrastructures, and that for these to be perceived more favourably, they need to change.

52% of respondents agreed or strongly agreed that collecting and analysing data can be good for society. This shows that whilst having some concerns about data uses, as seen in the responses listed in the previous paragraph, people can also recognise the benefits of data uses. However, only 12% strongly agreed with this statement, and 34% of respondents neither agreed nor disagreed, which is a larger percentage than for any of the other statements. The low fraction of respondents strongly agreeing with this statement is indicative of concern.

Figure 8: Please indicate how strongly you agree or disagree with the following statements.



6.2. Demographic differences in attitudes to data uses

There are few differences between groups with respect to attitudes to data uses. Exceptions relate to disability, gender and qualifications, but all differences were small.

There were differences in attitudes between disabled people and people who are not disabled. In most cases the differences were small, with two exceptions: disabled people were more positive about the statement 'I support the re-use and sharing of health data for research purposes', with 60% agreeing with the statement, compared with 54% of people who are not disabled. They were also more concerned by commercial companies providing public services (70% compared to 65% respectively).

As with differences relating to disability, most gender differences in attitudes to data uses were small. However, women were consistently more likely to want to know who has access to data about them (84% of women, 80% of men), to want more control over how their personal data is used by organisations (85% of women, 79% of men), and to know where data about them is stored (84% of women, 76% of men).

Differences by education were also mostly small. With that said, people with more qualifications were much more likely to agree with the statement 'Collecting and analysing data can be good for society', with 59% of people with degrees agreeing compared with 38% of people with no qualifications. People with more qualifications are moderately more likely to support the re-use and sharing of data for health purposes, with 59% and 53% respectively.

In general, attitudes towards data practices were largely similar across the different groups we surveyed, and where there are differences between groups, they tend to be small. Research suggests that the impacts of data uses are experienced unevenly by different groups, with marginalised and disadvantaged groups more likely to be negatively affected (eg Eubanks 2018). Such differences were not strongly reflected in different groups' attitudes towards data uses in our survey, suggesting that other factors, such as knowledge and understanding, may play an important role.

6.3. Context is a defining factor in concerns about data uses

As noted above, we identified high levels of concern about data uses in responses to attitudinal questions. However, it is important to note that some data uses are more concerning than others, and there are important differences in attitudes depending on the type, context or purpose of data use. Context is a defining factor in concerns about data uses, such that it is not helpful to ask people about their attitudes to data uses out of context.

In Figure 8 above, there is more support for using health data for research purposes (55% agree or strongly agree with the statement 'I support the re-use and sharing of health data for research purposes') than there is for companies monitoring children's mobile phone use for well-being purposes (37% of respondents agree or strongly agree with the statement 'It's OK for companies to use software to monitor children's mobile phone use, if it's to support their well-being'). Indeed, the whole purpose of our Living With Data research is to understand how differences in context, type and purpose of data practice influences attitudes to them. Section (9) below, on perceptions of specific public sector data practices, discusses this point in greater detail.

7. Comparing concerns about data uses with other concerns

It's possible that, when asked their views on data uses, respondents express concern, yet in their everyday lives, they do not actually worry very much about data uses. To understand how concerns about data uses compare with other concerns, we asked respondents how concerned they were about two lists of issues and to identify the three items in each list about which they were most and least concerned. The items in the first list were broad concerns, derived from Ipsos Mori's Most Important Issue tracker, such as the economy, pandemic diseases and immigration, alongside 'data practices'. The items in the second list were narrower concerns, such as funding for the NHS or anti-social behaviour and local crime. In this list, we included a number of specific data practices, such as 'automated technologies being used to try to change people's behaviour', 'personal health data being used to manage Covid-19' and 'data being used in unfair ways'. In each case, items were presented to respondents in a random order.

As can be seen in Figure 9, compared to other broad concerns, concern about data uses was ranked 8th out of 15, less concerning than health and the environment, for example, but more concerning than potentially divisive issues like Brexit and racism. Figure 10 shows that the smallest number of respondents listed it among the three items they were most concerned about (6%), and a large number listed it as one of the three items they were least concerned about (28%). This combination suggests that while a majority of people are concerned about data uses, it is rarely among people's main concerns. Similar fractions of people state they are either 'not very concerned' or 'not at all concerned' about data practices, and put data practices among the three items they're least concerned about. Coming at the end of a survey about attitudes to data uses, it is possible that this ranking may be higher than it would be if the question was asked at the beginning of the survey, or in a different context, and this may account for this tension. It is also worth highlighting that some of the items ranked below data practices in figure 9, such as immigration and Brexit, have relatively larger fractions of people selecting "Very concerned" than the equivalent for data practices; it is unsurprising that these items then have more people selecting them as among their top three concerns.

Figure 9: In general in your daily life, how concerned are you about each of the following?

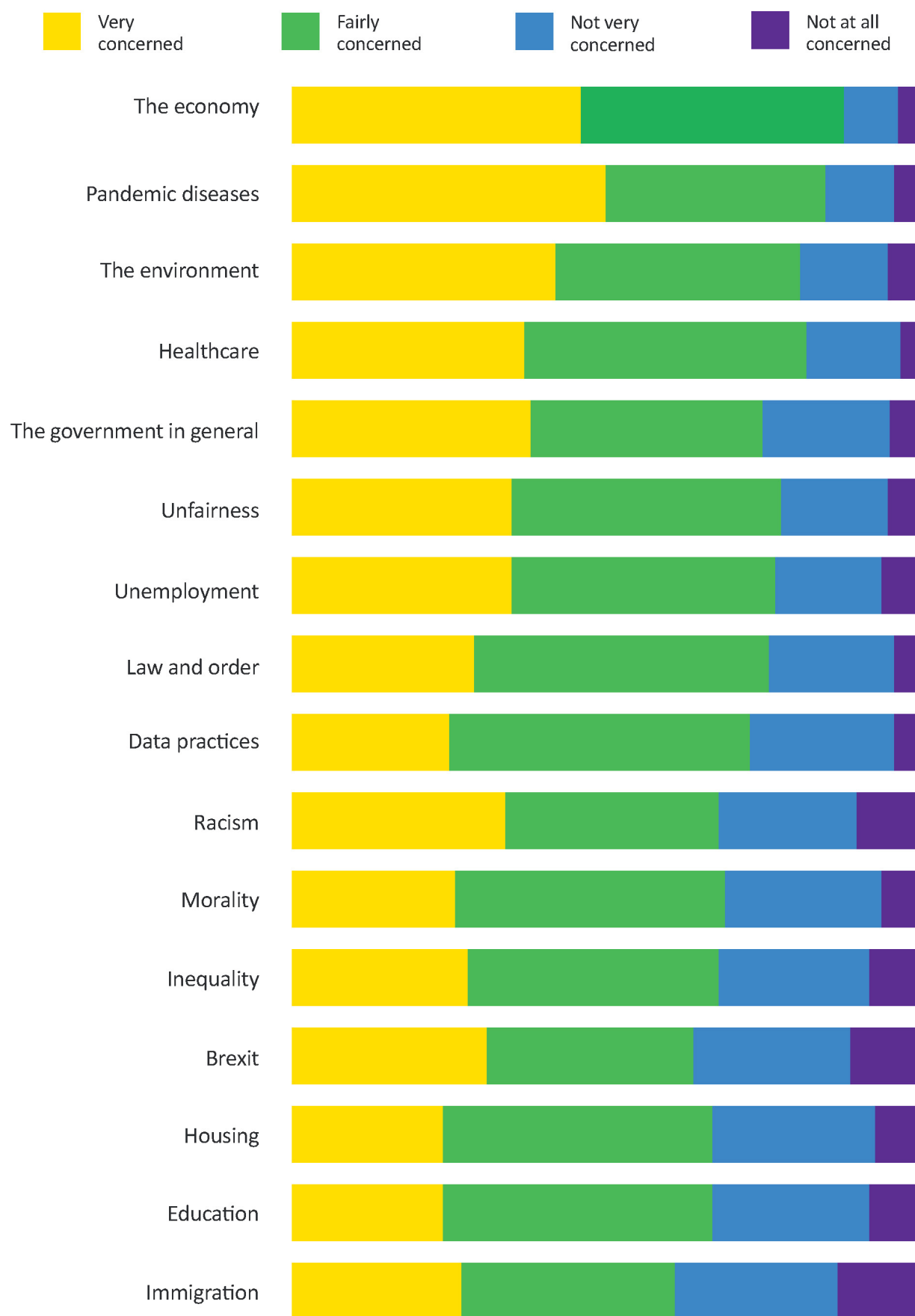
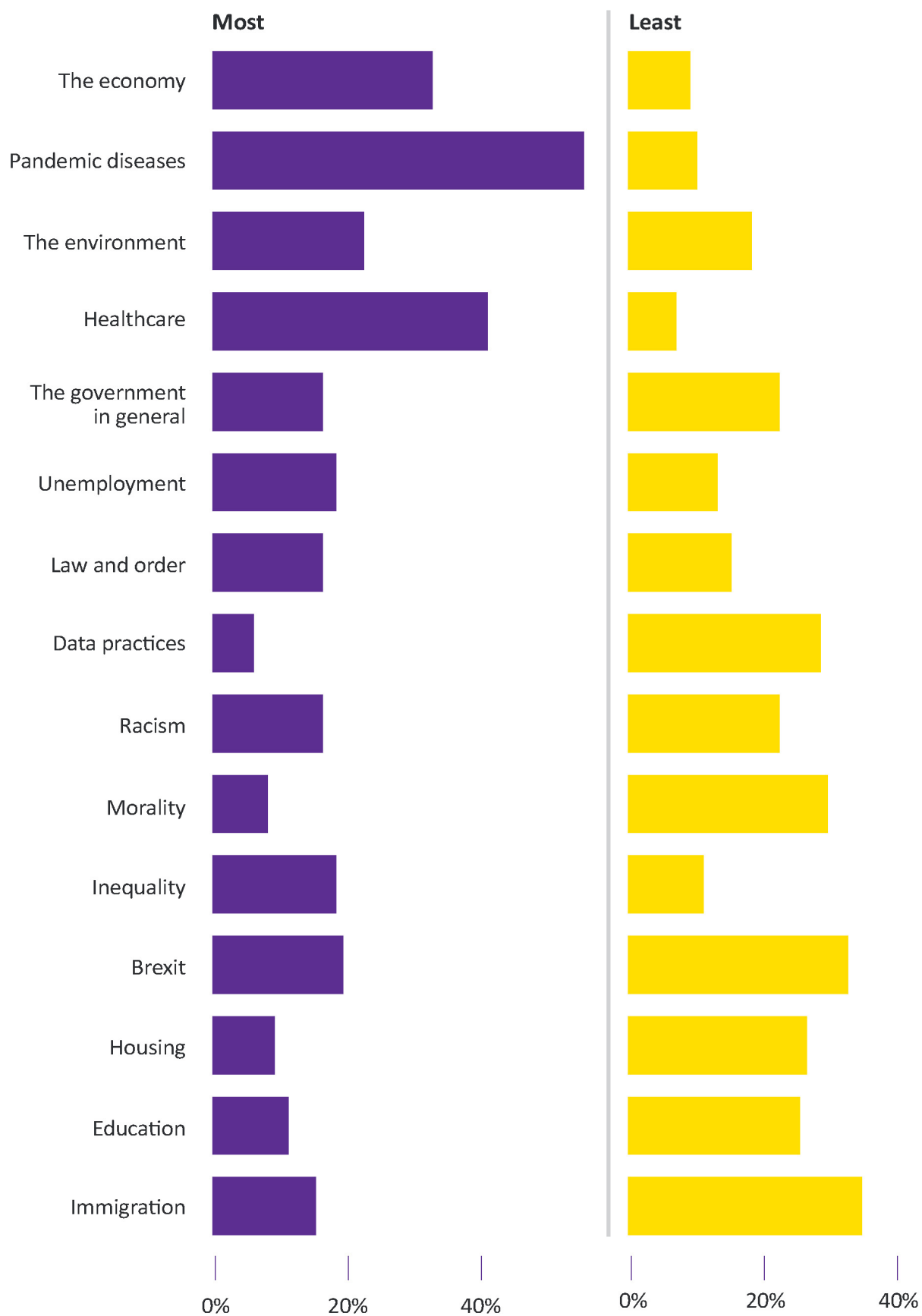


Figure 10: We'd like you to tell us which three things you're most concerned about, and which three things you're least concerned about.



However, a different picture emerged in responses to questions about more specific concerns. Here, only the economic costs of Covid-19 and funding for the NHS ranked more highly than data being used in unfair ways, as seen in Figure 11. Specific data-related concerns, namely a) commercial companies profiting from personal data, b) organisations tracking when, where and how people log on and c) automated technologies being used to try to change people's behaviours were concerning, ranked the 5th, 6th and 7th biggest specific concerns from a list of 13. Personal data being used to manage Covid-19 was the least concerning from the list of specific concerns. And yet, list items associated with data uses were rarely in respondents' top three concerns, whereas the economic costs of Covid-19 and funding for the NHS were both selected by a majority of respondents.

From these findings we can conclude again that the context, type and purpose of data use are important in determining people's thoughts and feelings about them. We can also see that when asked if they are concerned about the rather opaque issue of 'data practices,' people are moderately concerned, but this is rarely their main concern. In contrast, asking people their views on more specific aspects of data use helps to unveil different attitudes to different practices, some of considerable concern, such as data being used in unfair ways, some of very little concern, such as personal health data being used to manage Covid-19). Data from free text fields elsewhere in the survey indicate that although respondents support gathering data to manage Covid-19 in principle, they have concerns about how data is shared and with whom. We say more about this below.

Figure 11: In general in your daily life, how concerned are you about each of the following?

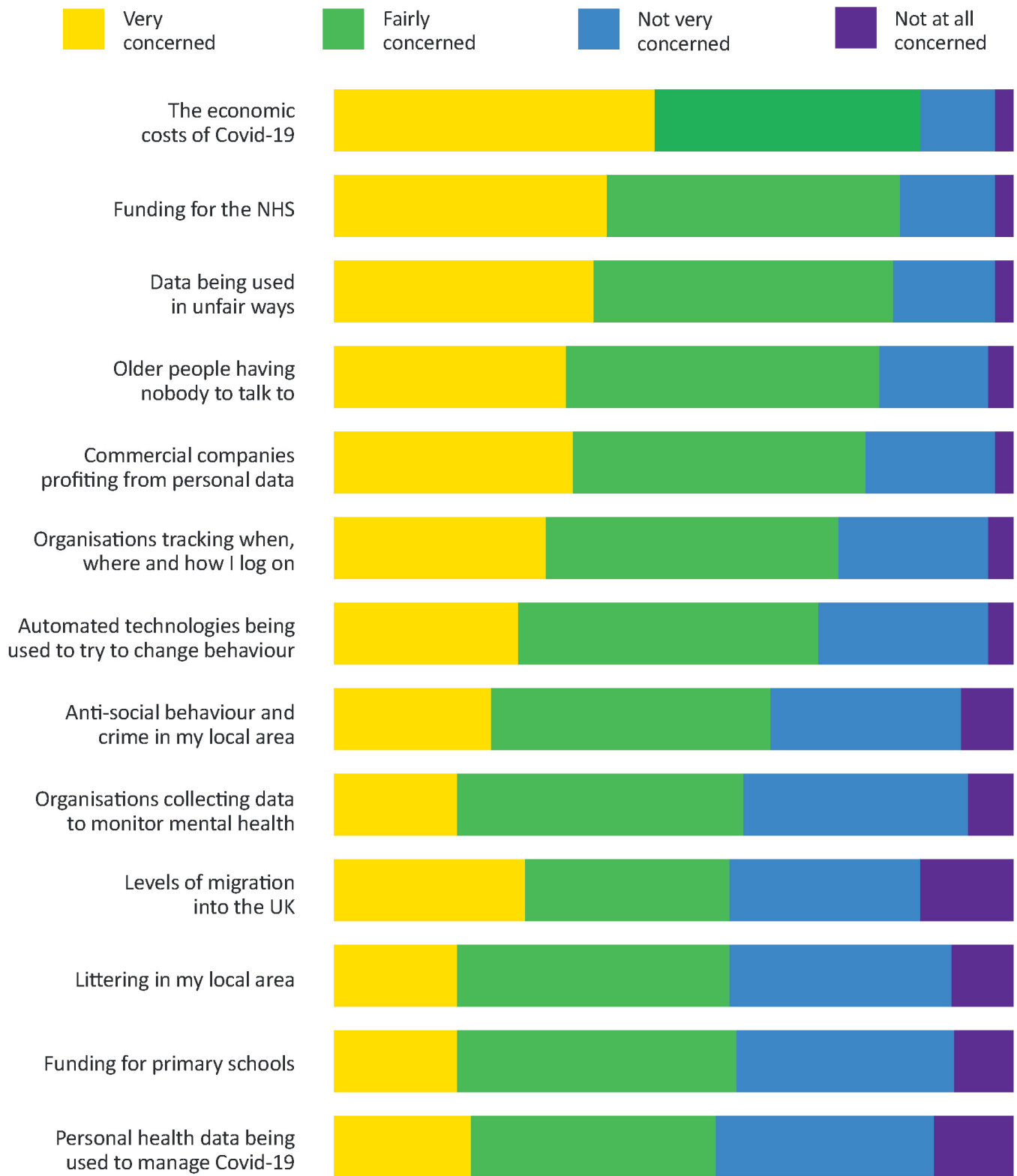
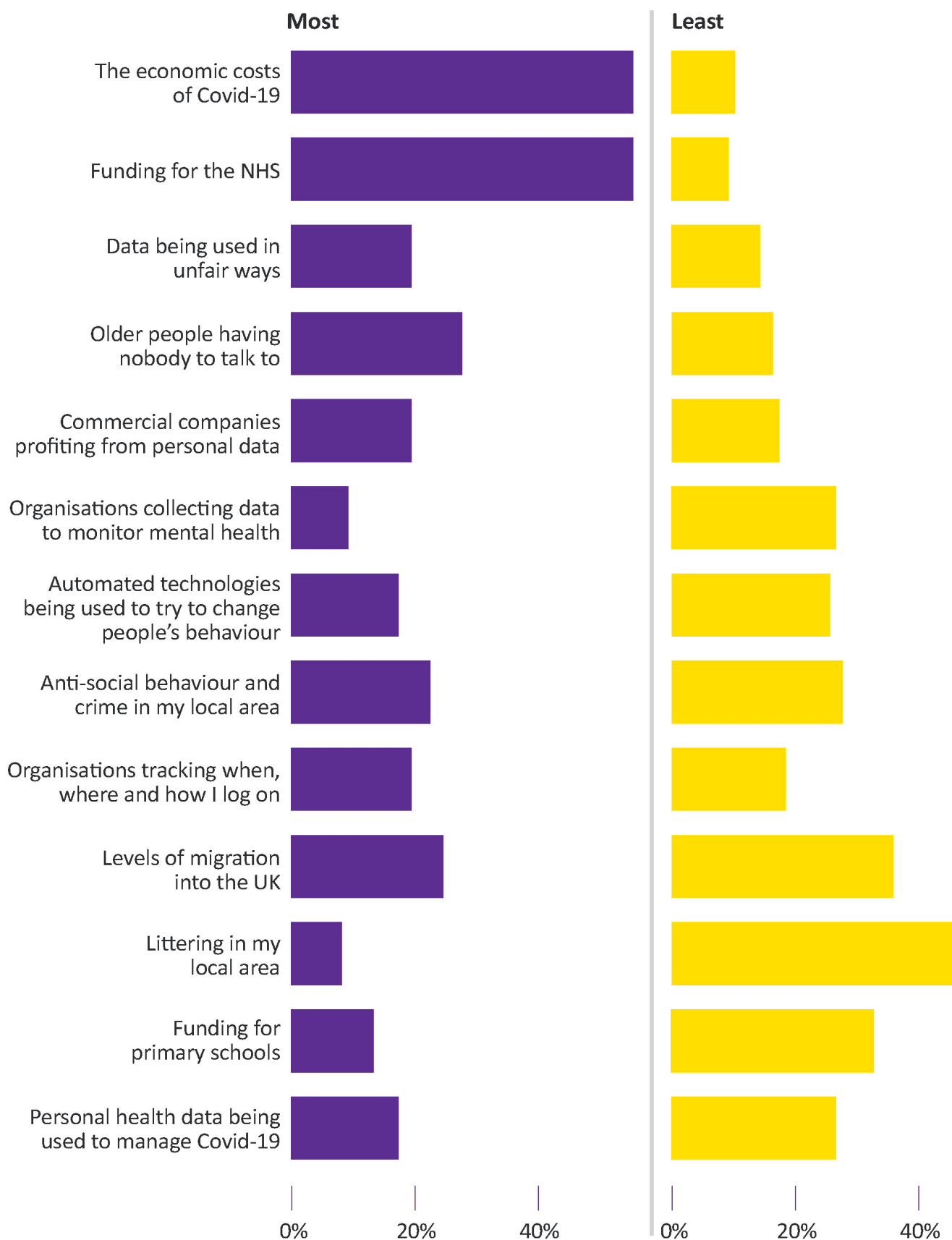


Figure 12: We'd like you to tell us which three things you're most concerned about, and which three things you're least concerned about.



7.1. Concerns and demographic differences

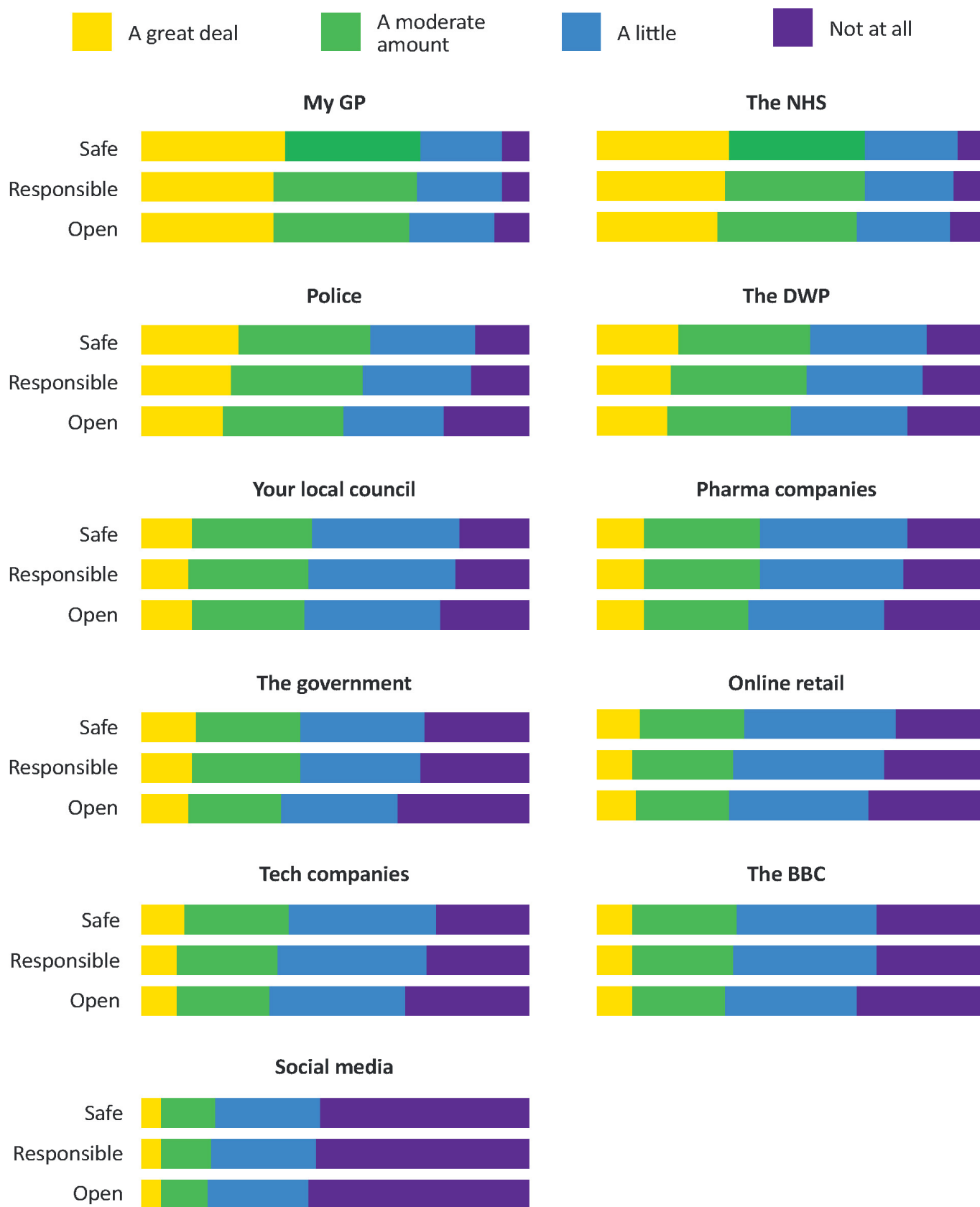
People of colour, LGBTQ+ people, and people in receipt of Universal Credit were more concerned about data uses than other groups. Differences were in most cases larger for concerns relating to the specific data uses in the second list than general data uses in the first list, though differences were never very large. Younger people (50% vs 24%), women (40% vs 27%) and people of colour (65% vs 28%) are more concerned about racism, while older people (38% vs 17%) and white people (29% vs 25%) are more concerned about immigration. Higher-income people (20% vs 29%) and people with higher qualifications (19% vs 39%) are also pronouncedly less concerned about immigration. In addition to this, the other very large difference is that people with disabilities are significantly more concerned about healthcare (44% vs 34%).

There are no differences by age relating to concern about general data uses. People of colour were more concerned about data being used in unfair ways and personal health data being used to manage Covid-19, and people in receipt of Universal Credit were more concerned about data being used in unfair ways. These differences were not large; for example, 27% of people in receipt of Universal Credit had data being used in unfair ways as one of their top three concerns, compared with 18% of people not in receipt of Universal Credit. LGBTQ+ people were more concerned about general and specific data practices. There were no differences by gender, qualification, disability, or income.

8. Trust in institutions' data uses

Respondents answered questions about their trust in a number of different sectors (such as social media, online line, the pharmaceutical industry) and institutions (such as the police, DWP, NHS and BBC) to a) keep their data safe, b) gather and analyse data about them in responsible ways, and c) be open and transparent about what they do with data. We asked these questions in order to gauge whether trust varies across institutions and data uses, and whether trust in organisations and sectors in general influences attitudes to those same organisations' or sectors' specific data uses. We included the BBC, DWP and NHS as their data uses were the focus of the section of the survey about attitudes to specific data uses. Overall responses can be seen in Figure 13 below.

Figure 13: How much do you trust <organisation> to: keep data about you safe?; gather and analyse data about you in responsible ways?; be open and transparent about what they do with data about you?



We found that respondents said they trusted public health care professionals – specifically their GPs and the NHS – most with their data, and they trust media and social media companies the least. Low levels of trust in the BBC’s data uses were noteworthy. In comments in free text fields in the section of the survey that explored attitudes to specific public sector data uses (discussed in Section 9), respondents also indicated a lack of trust in the BBC as an organisation. Also in these free text comments, a nuanced picture of trust in relation to a specific NHS initiative, the NHS Covid Data Store, emerged with similar numbers implying they had trust in the data process as those who implied they did not. Comments showed that respondents supported the aims of the data store, but they were concerned about the involvement of commercial companies (see Section 9.2 for further discussion).

Levels of trust expressed by respondents were consistent across the three data uses that we asked about (keeping data safe, gathering and analysing data in responsible ways, and being open and transparent about what is done with data). For example, 5% of respondents trusted social media companies a great deal with regard to all three of these activities. 34% of respondents said they trusted the NHS a great deal to keep their data safe, 33% trusted the NHS a great deal to gather and analyse data in responsible ways and 31% trusted the NHS to be open and transparent about what they do with people’s data. This consistency in degrees of trust across the three data uses suggests that respondents’ trust in sectors and institutions strongly influences their trust in the same sectors’ and institutions’ data uses. In other words, here we find again that context is a defining factor when it comes to attitudes to data uses.

Most differences across demographic groups related to age. 74% of those 65 and older trusted their GP to be open, 81% to be responsible, and 80% to be safe in their data uses, compared with 65%, 62%, and 62% respectively for the youngest group. Older people had more trust in the police in relation to using data responsibly and keeping it safe. Older people were also more trusting of the DWP – for example, 61% trust the DWP to keep their data safe, compared with 45% of the youngest group, and significantly less trusting of social media companies, 8% compared with 22% of the youngest group.

Responses associated with income were less straightforward. The highest-paid respondents were more likely to give more extreme responses, stating that they trusted institutions ‘not at all’ or ‘a great deal’. Higher-paid respondents were more trusting of the police and of the DWP, with 60% of the highest-paid group trusting the DWP to be open about data uses, compared with 50% of the lowest-paid group. However, people claiming Universal Credit had more trust in the DWP’s data uses than people not claiming Universal Credit: again with openness as a comparison, 55% compared with 49%.

While White people were most likely to say they trusted the police ‘a great deal’ (28% to keep data safe), Asian people were likely to say they trusted the police ‘a moderate amount’ (44%). Black people and people in mixed and other ethnic groups were least likely to trust the police, with a third saying they did not trust the police to be open about their data uses at all, compared with 13% of White people. White people were also less likely to trust government and social media companies’ data uses (for example, 57% had no trust at all in social media companies to behave responsibly with their data, compared with 45% of Asian people and 38% of Black people). LGBTQ+ people were less likely to trust their GP and the DWP than heterosexual respondents (30% compared to 35%, and 16% compared to 19% respectively). Disabled people were more likely to trust their GP and the NHS in general (38% compared to 33%, and 36% compared to 30% respectively), than respondents who did not report a disability.

9. Attitudes to specific public sector data uses

We asked respondents about their attitudes to specific public sector data uses, in order to understand public perceptions of public sector, everyday, real-world data uses in specific domains. Respondents were randomly allocated to one of three sets of questions, about current or potential future data uses of the BBC, DWP or NHS. Questions about each data use differed, because the uses themselves differ. Each set of questions concluded with questions about how much respondents felt they understood the data uses (to explore how understandable specific data uses are) and how surprised they were to read about them (to explore respondents' expectations about these public sector organisations' engagement in such practices). Respondents were also given the option of providing more detail in a free text field about why they would/would not use a data-driven service or why they were/were not comfortable with a data use.

The BBC and DWP were partners on our *Living With Data* project and our contacts at these organisations selected data uses on which to focus. The BBC examples were experiments with 'personal data stores' (PDS), where a person's data is stored in a secure place that belongs to and is controlled by them, rather than being stored and controlled by organisations like the BBC, or other media companies like Netflix. The DWP selected 'Confirm Your Identity', on which it was experimenting with different data-driven ways of confirming that claimants are who they say they are. We did not partner with the NHS for this research; rather, we selected the NHS Covid Data Store as a focus for our survey, which was carried out during the Covid-19 pandemic.

To understand BBC and DWP data uses that were the focus of our research, we adapted some elements of Bates et al's (2016) 'data journeys' methodology, which the authors describe as a methodology for capturing how data is constituted and how it moves between different sites of practice. For our research, we analysed public facing and organisational documents that were made available to us, and we interviewed key staff working on relevant projects, in order to develop detailed descriptive accounts of these specific data uses. We used these accounts as a basis for our survey questions. The NHS Covid Data Store account and questions were based on publicly available documentation. We partnered with BBC and DWP in our research to enable access to some of their data uses, and in order to be able to share and discuss the implications of our findings with them. Both organisations chose examples of data uses that they were willing to have scrutinised by our research team. Different examples would have resulted in different findings – more controversial data uses may have elicited more concern from respondents, for example. However, it is noteworthy that some aspects of the chosen examples did elicit concern, from which we might conclude that many data uses have elements that elicit concern.

9.1. BBC data uses

The BBC data practices that we focused on were BBC Box and BBC Own It, experiments with personal control over personal data in the form of 'personal data stores,' or PDS. In a PDS, people can often edit or add to the stored data, or they can choose to share their data with an organisation like the BBC, in exchange for recommendations, for example of programmes to watch or listen to. Respondents answered questions about BBC Box and BBC Own It having read descriptions of them. For BBC Box, respondents answered questions about how likely they would be to use it, either in the cloud or as a physical device. For Own It, they read a shorter initial description, and then were presented with a series of further details. For each detail, respondents were asked to record how they felt about it on a slider, with options from 'Not at all comfortable' at one end to 'Very comfortable' at the other.

BBC Box

The text that respondents read about BBC Box was as follows:

BBC Box pulls together data about what you watch or listen to (for example on the BBC or Spotify) in one single place. BBC Box then creates a user profile identifying your likes and interests, based on this data and quizzes that you have completed. You have the power to edit your profile. This data and your profile are stored securely in your personal data store. You can then choose to share your edited user profile with the BBC or other services. If you share your profile, you will receive recommendations, not only about what you might like to watch or listen to, but also about things to do and places to visit.

In its initial version, the BBC Box was a physical object. Data was kept on the physical device and your user profile was created there. In its second version, the BBC Box is cloud-based – that is, data is stored and your user profile is created securely online. BBC Box is accessed via an app on a user's device.

To get recommendations - for example of TV programmes to watch, concerts you might want to go to, or holiday destinations - you would need to share your user profile with the BBC or another organisation. BBC Box is in testing, not currently available for use.

Figure 14 shows how likely respondents felt they would be to use BBC Box as a physical device, if it became available for use, and Figure 14 shows how likely respondents felt they would be to use it in the cloud if it became available for use.

Figure 14: How likely would you be to use BBC Box as a physical device?

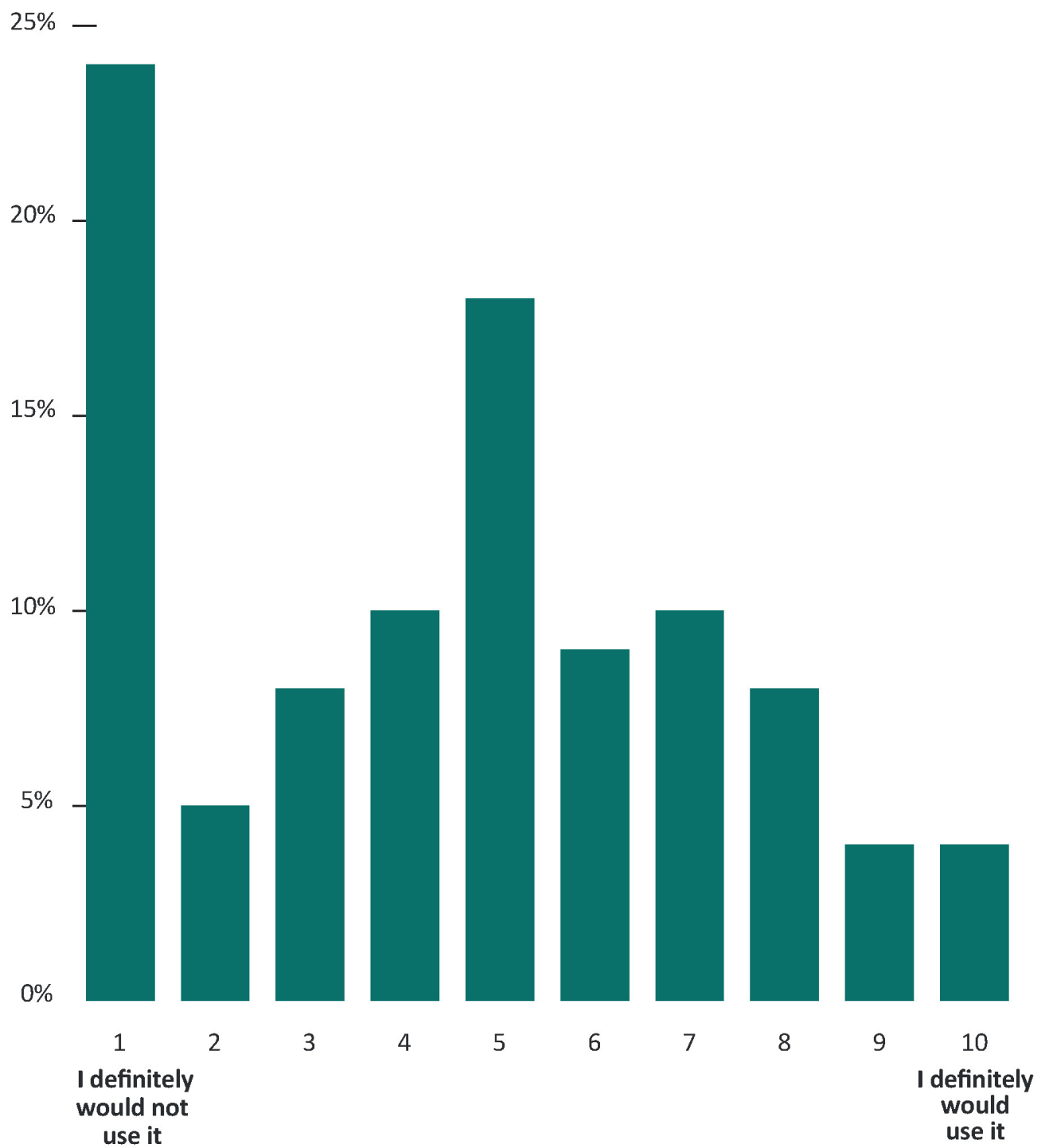
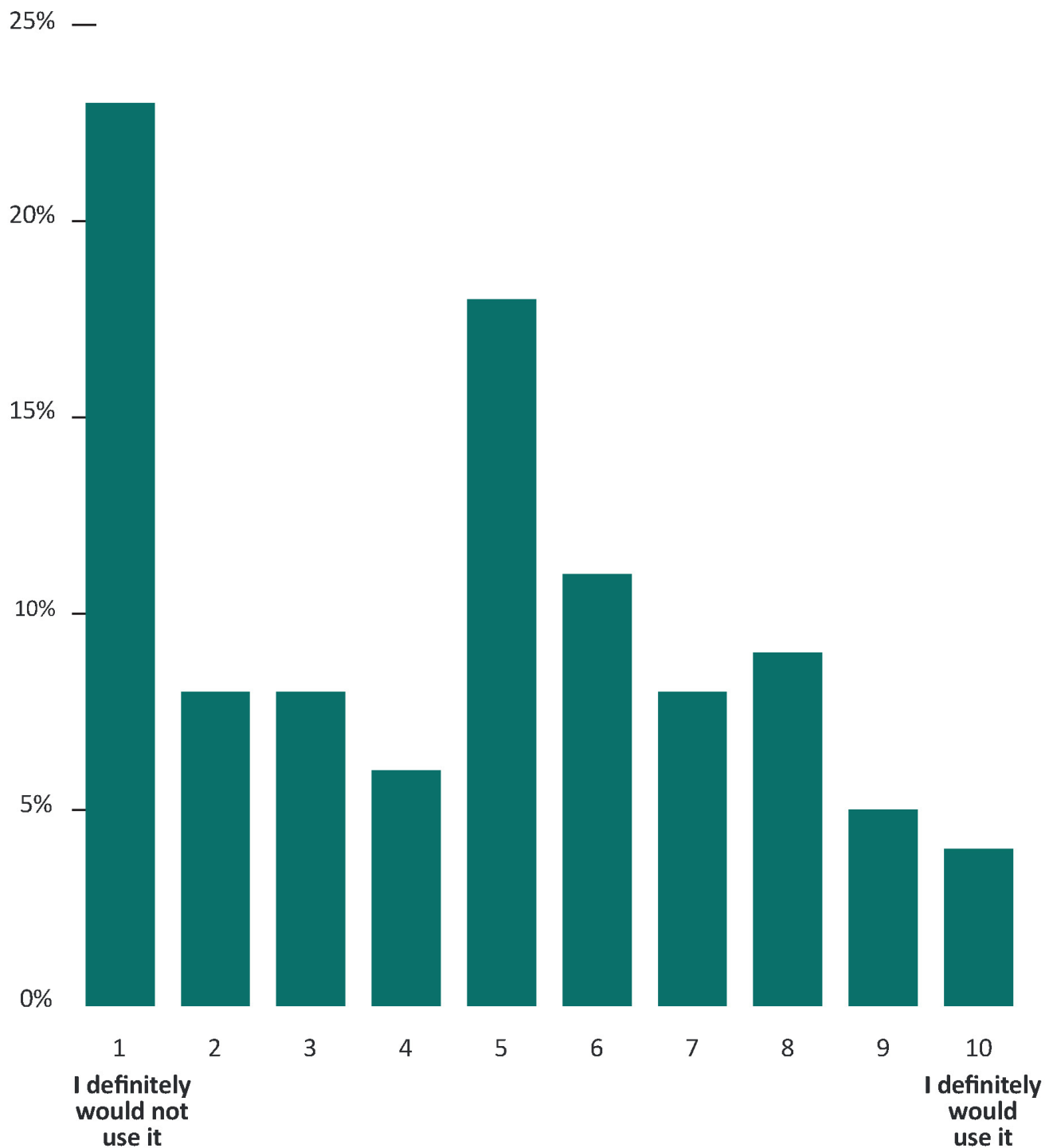


Figure 15: How likely would you be to use BBC Box in the cloud?

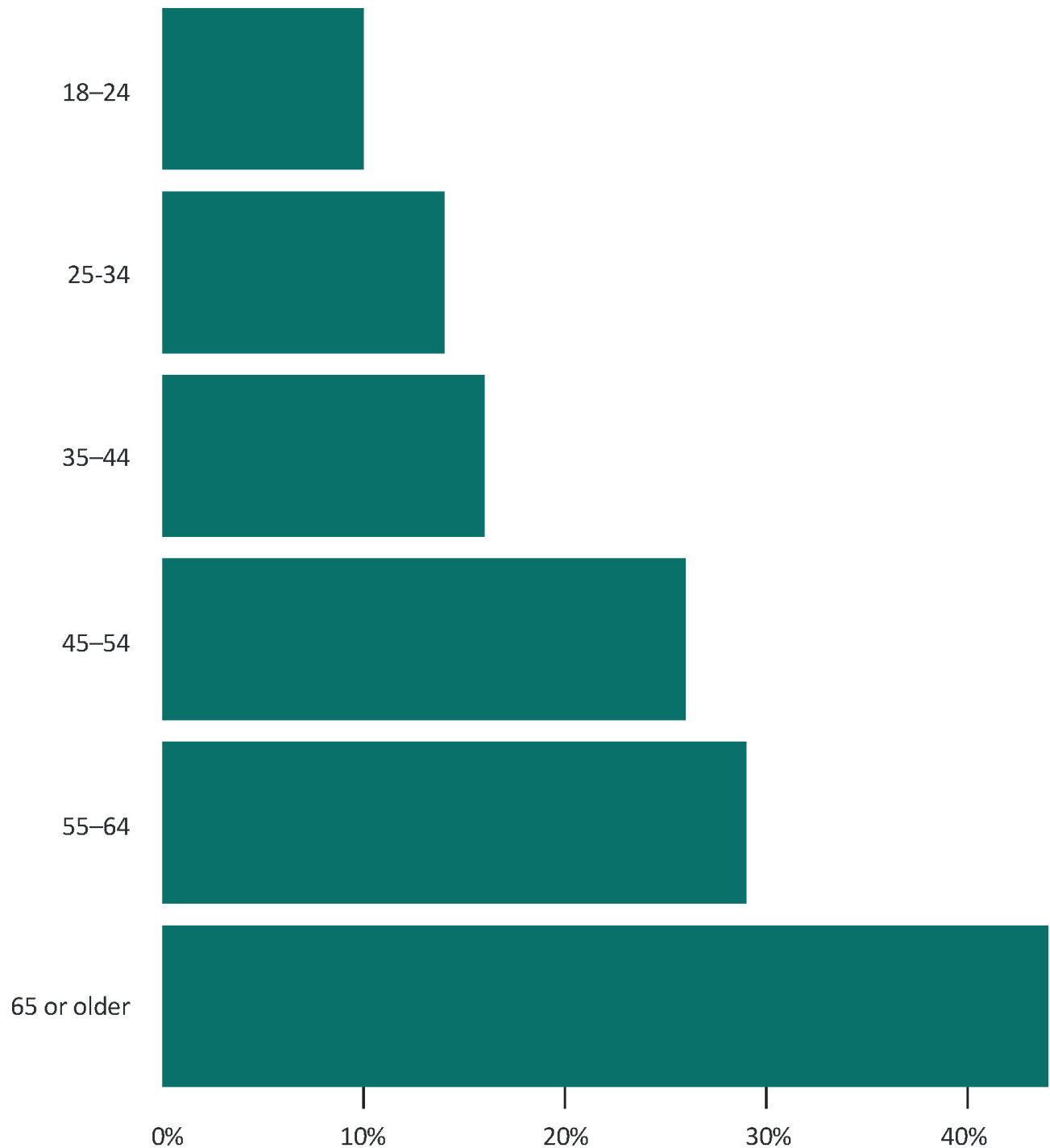


As can be seen, there was little difference in responses across the two modes of possible BBC Box use, suggesting that mode of use was not important to them. The figures also show that the most common response was that respondents definitely would not use BBC Box. Just 4% of people said they definitely would use it in each case. Responses in free text fields suggest that this is because of: a) a lack of interest

or perceived need in the product; b) low levels of trust in the BBC; and c) wanting more information on what the BBC Box does. 16% of free text comments about BBC Box indicated lack of interest in and need for BBC Box compared with 2% of comments about the NHS Covid Data Store and 3% about the DWP identity verification process. Lack of trust in the organisation undertaking the data practice was also mentioned more frequently in relation to BBC Box than in relation to the other data uses examined in this section of the survey. This can also be seen in data from elsewhere in the survey: of the people who said they trusted the BBC a great deal to be open with what they do with users' data, 21% said they would definitely use a physical BBC Box, compared with just 3% of those who said they did not trust the BBC at all on this same issue. Finally, respondents indicated that they 'needed to know more' about the BBC Box in 5% of the free text fields compared to 0.4% for DWP identity verification. The second most common response to the question about whether respondents would use BBC Box was in the middle of 'definitely would not use it' and 'definitely would use it', which could also be interpreted as a need for more information.

In terms of differences across groups, younger people were much more positive about using BBC Box. Although small numbers of younger respondents said that they would definitely use BBC Box, consistent with all respondents, there were pronounced differences in those who said they definitely would not use it, which increase with age for both the physical and cloud variations. This is illustrated in Figure 16, which shows the percentages of people in each age group who reported that they definitely would not use BBC Box in the cloud.

Figure 16: How likely would you be to use BBC Box in the cloud? Percentage reporting “I definitely would not use it”, by age group



There were also differences by qualifications. Although the most common response for all groups is ‘I definitely would not use it’, people with higher qualifications were more positive about BBC Box both as a physical device and in the cloud. There were also differences by ethnic group: white people had more

negative responses to BBC Box, while people of colour had more positive responses to it³. Finally, there were some smaller differences by gender, with women having slightly more positive attitudes towards BBC Box than men.

BBC Own It

Next, respondents were presented with some information about BBC Own It. They were told that Own It, currently available for use (unlike BBC Box, which was an experiment not in use at the time of the survey), is a free app designed by the BBC to support, help and advise children when they use their phones to chat and explore the online world, without adult supervision. They were told that Own It consists of a custom keyboard and companion application. The keyboard becomes the default keyboard for all apps used on the phone onto which the Own It app has been downloaded. After a child has downloaded Own It, what they type into the phone and the pages they visit are processed by the app. Own It assesses children's phone use to provide 'in-the-moment support' (in the form of alerts or recommendations) for issues that children may experience, by scanning what they do on their phones.

Respondents were then asked to express their degree of comfort with a series of statements about Own It by dragging a slider which ranged from 'Not at all comfortable' to 'Very comfortable'. The statements were as follows:

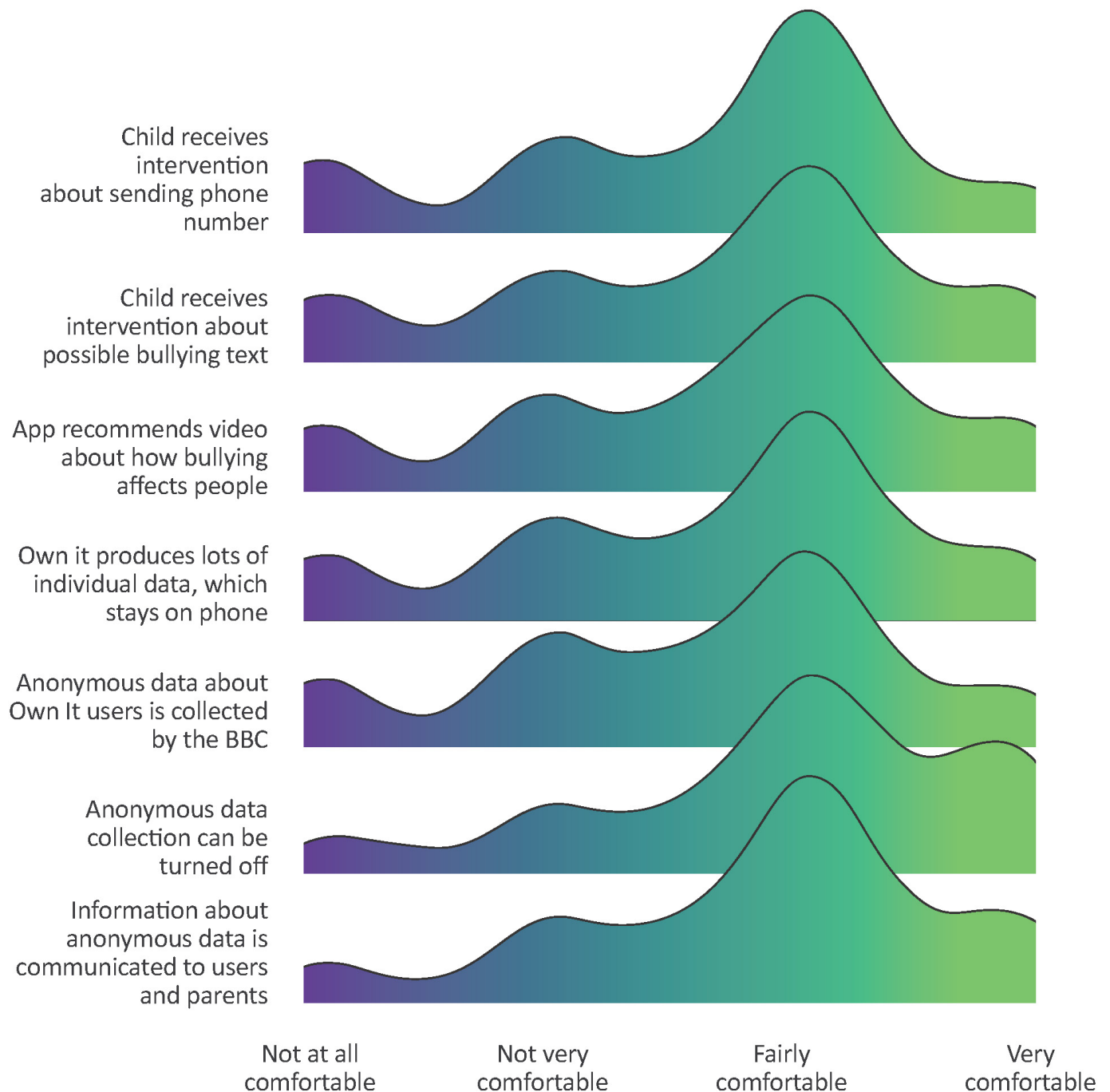
- A child has the Own It app installed on her phone. She is about to send her phone number to a contact on a social media platform popular with children. Own It intervenes and says 'are you sure you want to send your phone number?'
- A child is typing a message to a friend and receives an alert from Own It asking him whether he is sure he wants to send the message. This is because the Own It app has scanned the text of the message and identified words which suggest the child might be bullying his friend.
- After Own It sends this alert, it recommends that the child watches a video about how bullying affects people.
- Using Own It produces data about: how much time a child has spent on their phone, how the phone is used, pages visited, how the child feels (gathered from quizzes and questionnaires within the app), 'in-the-moment support', the child's feedback on the usefulness of that support. The BBC will not see any data connected to an individual child, and all identifiable data stays on the phone.
- Anonymous data about Own It users is collected by the BBC. For example, 'the emotion anger was identified 250 times among all children', or '30 out of 200 children did not find this content recommendation useful'. This data is used by app developers to help them improve how the app works.
- Users can turn this feature off at any point.
- Information about anonymous data collection and the option to turn this feature off is communicated to Own It users and their parents. (This information can be found in Information for Parents section of the app.)

Responses to these statements are shown in Figure 16. These curves denote the slider positions for all respondents. If half of respondents had dragged the slider all the way to the left of the scale (= not at all comfortable), and half had dragged the slider to the centre (= between not very and fairly comfortable), the curves would have significant peaks at the far left and in the middle. Where the curve is high, a large

³ We group people of colour together here because of the smaller numbers of respondents who answered questions in this section of the survey.

fraction of respondents dragged the slider to that point; where the curve is low, a small fraction of respondents dragged the slider to around that point.

Figure 17: Responses to statements about Own It



Most respondents were fairly or very comfortable with the data uses described, although there are some differences. Respondents were least comfortable with anonymous data about Own It users being collected by the BBC and most comfortable with the fact that this feature could be turned off.

Following these questions, respondents were informed that the BBC often collaborates with university researchers, sharing anonymous data with them to help evaluate and develop products like Own It. Respondents were told that the BBC is considering making more detailed Own It data available so researchers can better understand children's mental health. They were asked what data, in their opinion, should be shared. The options and results are presented below. The data sharing models were:

1. No sharing of any data – what's on the app stays securely on the mobile device (43% of respondents chose this option)
2. Sharing anonymised data (for example, 'on a given day, the average user spent one hour on the app', or 'there were 150 reports of children using the app reporting feeling angry' (27%)
3. Sharing anonymised, group level data (for example, 'of all children who completed a personality quiz and were categorised as having a particular personality, 20% reported often feeling angry') (22%)
4. Sharing anonymised, individual level data (for example, 'userID1234 reported feeling angry four times this week, gave these answers to a personality quiz and watched this video') (8%).

The majority free text field comments about BBC Own It that expressed concern about data sharing identified the lack of parental control or data sharing with parents as an issue.

We then asked respondents how comfortable they would be with a child they know using Own It. 59% said they would be very comfortable or somewhat comfortable. The most common response was somewhat comfortable (45%) with a child they knew using the app. Roughly similar numbers would feel very comfortable and not at all comfortable (14% and 16% respectively) with a child they know using Own It. This suggests that the majority of respondents are towards the middle of the scale, neither very comfortable nor very uncomfortable with a child they knew using the app.

By contrast to the responses to questions about BBC Box, the responses to questions about BBC Own It were more consistent across demographic groups. The exceptions to this were with age and ethnic group. In general, older people had more negative responses to the different dimensions of Own It, although these differences were not enormous. Similarly, people of colour were broadly more positive about Own It than White people were, although again differences were relatively small.

In response to the question about preferred data sharing models for Own It data, there were some differences. Younger people were more comfortable with group-level and anonymised sharing, with 36 and 38% of the age groups 18-24 and 35-44 stating they were most comfortable with no sharing, compared to 52% of the oldest group. White people were also more likely to feel most comfortable with no sharing, with 46% of White people choosing this option compared with 34% of people of colour. People with fewer qualifications were similarly likely to feel most comfortable with no sharing; the group that most preferred no sharing was the group with vocational qualifications.

In terms of people's comfort with Own It, younger people, people of colour and more highly-paid people were more comfortable with Own It than other respondents. People with intermediate qualifications (those taken at age 16 and 18) were the most comfortable, with people with fewer qualifications the least comfortable.

Finally, respondents were asked how surprised they were to learn about these BBC data practices, and how much they felt they understood them. Respondents were fairly or very surprised to learn about the BBC's experiments with BBC Box and Own It (76% across the two options). Only 5% of respondents were

not at all surprised, suggesting that the vast majority of respondents had not imagined such products to be available or had not expected the BBC to engage in such data practices. Older people expressed more surprise, as did people who were more highly-paid and people with no qualifications: 37%, 36%, and 48% compared with an overall average of 28% across the entire sample.

In response to the question about how much respondents felt that they understood these data uses, most responded that they somewhat understood both BBC Box (69%) and Own It (68%). A slightly larger fraction responded that they fully understood Own It (19%) than that they didn't understand it (13%), whereas numbers of respondents giving these answers in relation to BBC Box were more consistent, at 15% and 16% respectively. There were some major differences in self-perceived understanding. Younger people, more highly-paid people, and people with more qualifications were more likely to say they understood these programmes fully and less likely to say they don't understand them (20 or 21%, compared with an average of 16%).

We think that when a survey respondent states that they understand a data use 'somewhat', this reflects a reasonable amount of understanding, given that data uses are often hard to understand and their relevance to people's lives is not always immediately obvious. As a team of researchers, we ourselves did not fully understand some of the data uses we researched in the survey, precisely for these reasons. That almost 70% of respondents felt they understood BBC Box and BBC Own It represents a good amount of understanding, in our view.

9.2. NHS Covid Data Store

One third of survey respondents answered questions about the NHS Covid data store. These respondents received a description of the ways in which data in the NHS Covid data store is collected, shared, and used by different organisations. As with other descriptive text included in the survey, this description was revised multiple times, in an effort to ensure that it was clear, accurate and balanced. This was challenging, because details about the NHS Covid data store were not fully available, and some contradictions existed in available textual information. Unlike for our example BBC and DWP data uses, NHS representatives were not partners on *Living With Data*, so we could not turn to them for clarification of uncertainties. Experts on patient and health data from Understanding Patient Data and medConfidential on the *Living With Data* advisory board helped us to ensure the description was as accurate as possible. The final version included in the survey is reproduced below:

NHS (National Health Service) organisations, including NHSX (responsible for digital changes to national healthcare), have been commissioned by the government to develop a national, secure data store to hold data in one place to help national organisations responsible for coordinating the Covid-19 response. Patient data related to Covid-19, from GPs and hospitals, as well as call data from 999 and the 111 coronavirus helpline, lab test data from Public Health England, and data from the Office for National Statistics (ONS) is checked by NHS England and then uploaded to the secure NHS data store. The data in the NHS data store does not identify any individual in its current state. However, it may be possible for somebody with the right skills to re-identify some of this data.

Commercial organisations, including Microsoft, Google, Amazon Web Services, Palantir Technologies and Faculty, are collaborating in the development of the NHS COVID-19 data store. They are providing cloud platform and infrastructure technologies for the functioning of the data store, data collection tools and data processing software, data modelling and support with data analytics capacity and capability. They have access to patient data to enable them to fulfil their roles.

NHSX says that when the pandemic subsides and the outbreak is contained, they will close the Covid-19 data store. They say that the data processing will stop and all data will be either destroyed or returned to the NHS once the public health emergency situation has ended.

What NHS organisations say about the Covid-19 data store has changed several times and information exists in different places. Patient data groups are concerned that not enough detail has been provided about contracts with partners to fully understand who has access to data, for what purposes and for how long. They are concerned that commercial companies may be able to use data from the data store to develop their technologies and potentially increase their profits.

Respondents answering questions about the NHS Covid data store were asked how comfortable they felt about their NHS patient data being added to the store, how surprised they were to read our description of it, and how much they felt they understood it. Responses to these three questions are shown in Figures 18—20.

Figure 18: How comfortable do you feel about your NHS patient data being added to the NHS Covid data store?

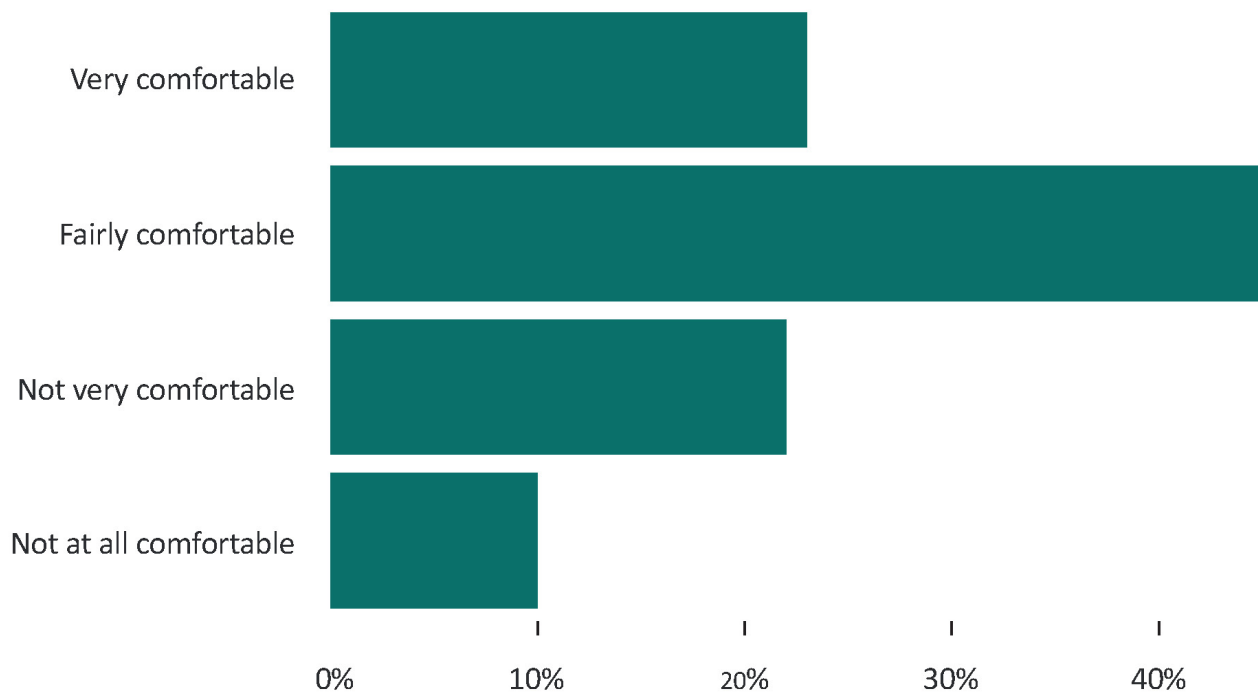


Figure 19: Are you surprised by the details of the NHS Covid data store?

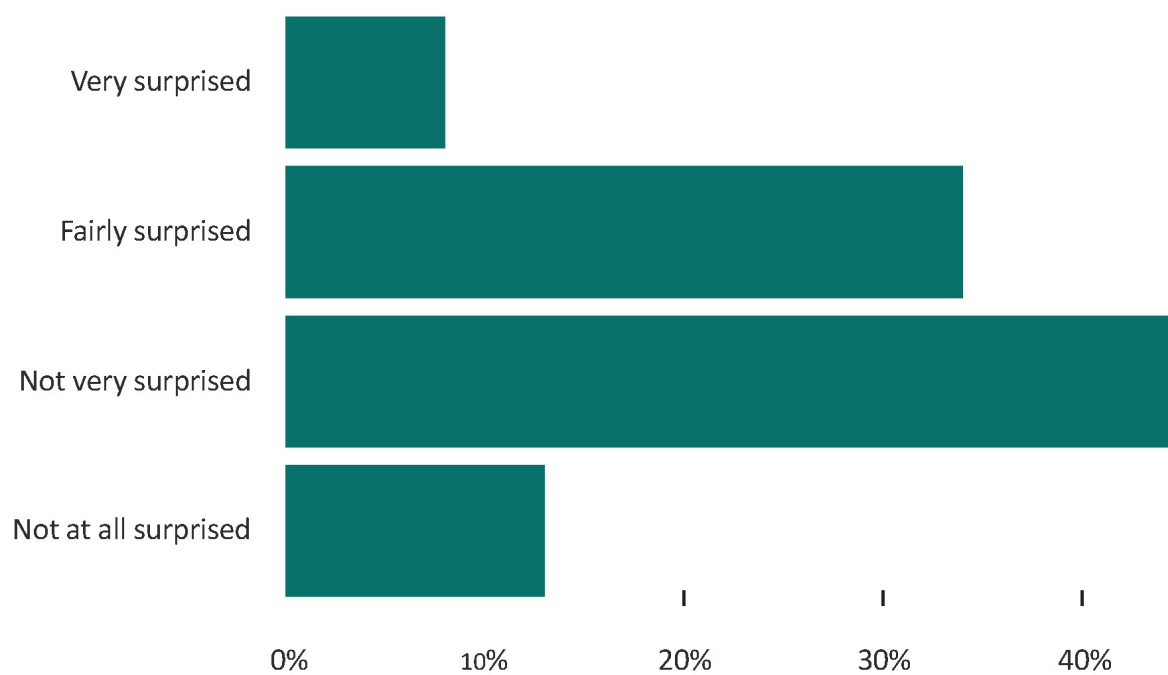
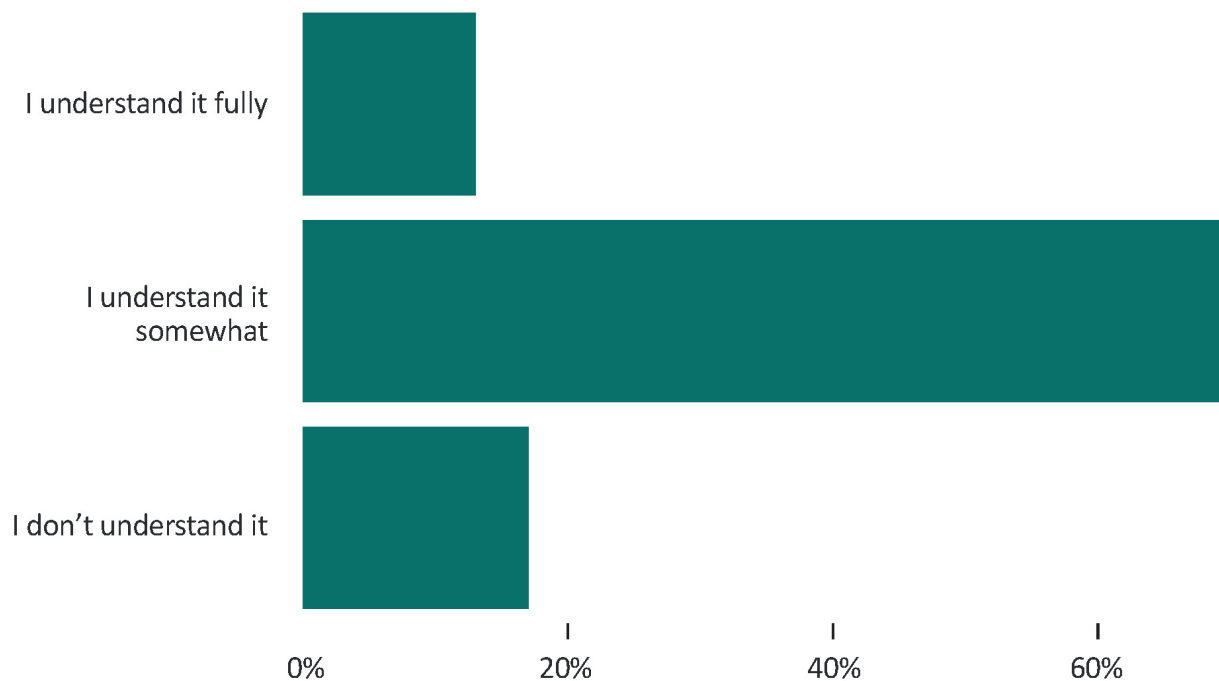


Figure 20: How much do you feel you understand the NHS Covid data store?



The majority of respondents were fairly or very comfortable about their NHS patient data being added to the NHS Covid data store (78%). Comments in free text fields suggest that this is because of support for its purpose. However, despite high levels of comfort, comments in free text fields about the NHS Covid data store revealed more concern about data sharing than with any of the other specific public sector data uses. More than half of these expressions of concern were about the involvement of commercial companies in the practice. Concerns were often expressed as imagined future scenarios, most of which were negative, and many of which involved commercial organisations profiting from, leaking, misusing or selling data in the future.

Although the majority of respondents were not surprised by the details of the data store, (58% were not very surprised (45%) or not at all surprised (13%)), it is noteworthy that a considerable proportion, 34%, were surprised. 70% of respondents felt that they understood the details of the NHS Covid data store somewhat. Remaining responses were distributed reasonably evenly across full understanding (13%) and not understanding (17%) this example.

As with BBC data practices, most groups had similar responses to the Covid data store, with some exceptions. There were differences in comfort with the Covid data store by ethnic group, gender, sexuality, and qualifications. People of colour were more likely to be comfortable with the Covid data store (73% very or fairly comfortable, compared with 66% for White people), as were men (73% compared with 64% for women). Broadly, people with higher qualifications were less likely to be comfortable with the Covid data store, and people with fewer qualifications were more likely to respond that they were not at all comfortable.

The only differences in surprise related to educational qualifications. As with the BBC data practices, people with higher qualifications were less likely to be surprised. People with no qualifications were the most likely to report being not at all surprised (24% vs 11% for people with degrees).

Other than ethnic group, these are the same as for the BBC data practices. Apart from the youngest group, younger people reported understanding the Covid data store more than older people did, and people of colour reported understanding more than White people did. More highly-paid groups also broadly report understanding more. The most pronounced differences between the lowest- and highest-paid respondents, with 29% of the highest-paid respondents stating that they understood the details of the NHS Covid data store fully, compared with 7% of the lowest-paid respondents. People with higher qualifications broadly report more understanding: the major difference is that people with no qualifications or qualifications taken around the age of 16 are the most likely to report that they don't understand: 29%, compared with 13% of people with degrees.

9.3. DWP Confirm Your Identity

Respondents answering questions about 'Confirm Your Identity' at the DWP were presented with this short introduction to the concept of identity confirmation which is at the heart of this project:

The Department for Work and Pensions (DWP) is the government department responsible for welfare, pensions, child maintenance and related policy. If someone needs to claim Universal Credit (a payment to help with living costs for people on low incomes), the DWP needs to confirm that the claimant is who they claim to be – this is known as confirming identity. Currently, most people take documents like a passport or driving license to a Job Centre to prove who they are. DWP is working on ways to make it possible to confirm identity online.

Respondents were then asked to record their responses to a series of statements on a slider, where at one end, the option was “Not at all comfortable” and at the other end, it was “Very comfortable”. The statements were:

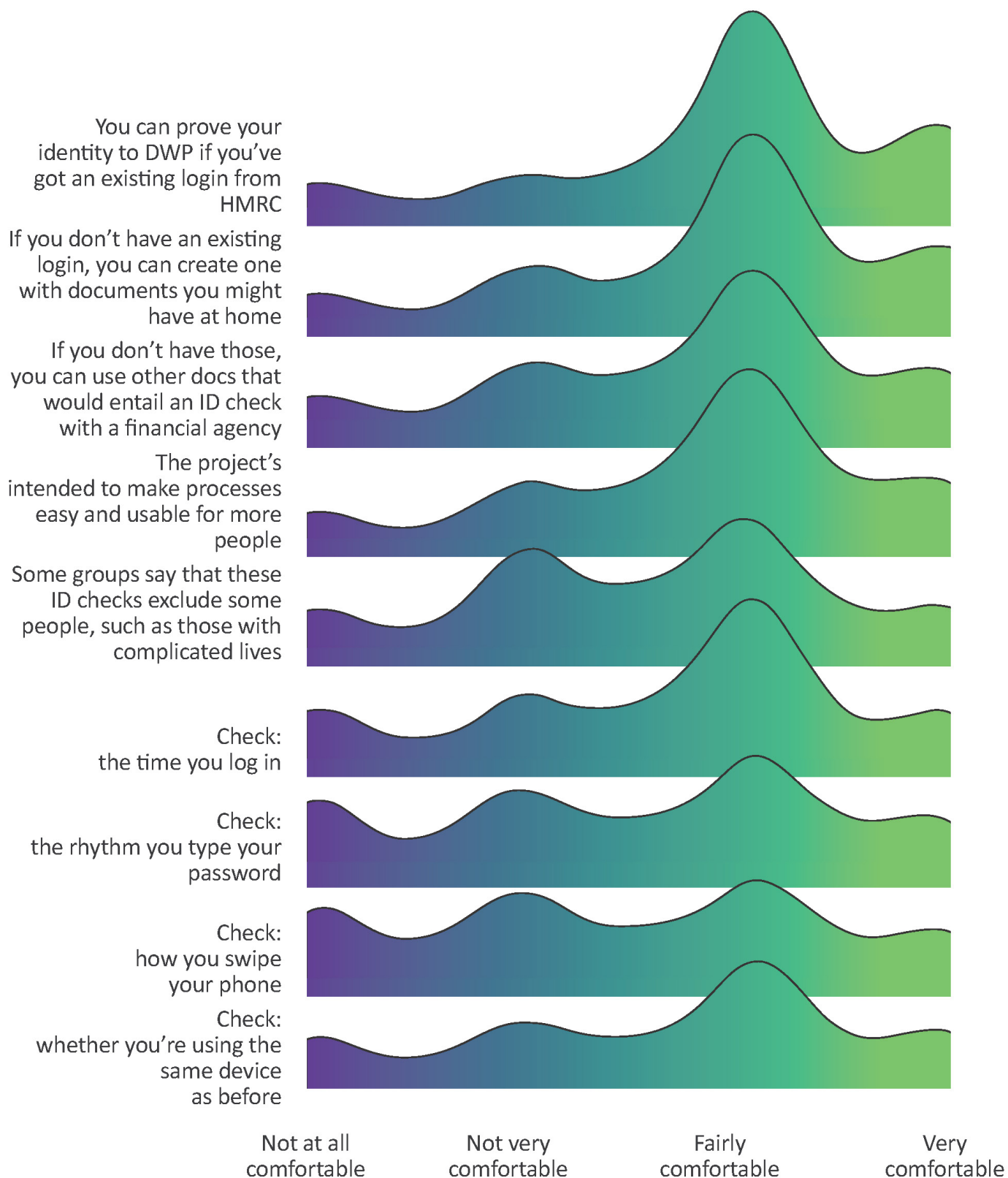
- *Imagine you are claiming Universal Credit from the DWP and you already have an online identity created by HMRC (the government department responsible for taxes and other financial matters) from a previous transaction with them. The DWP gives you the option to use an automated, secure system to get confirmation from HMRC that you have already proven your identity with HMRC. To do this you will need to login into your HMRC account when making your online Universal Credit claim and HMRC will check its records and send an automated confirmation back to the DWP. This way, you don't need to prove your identity again with the DWP.*
- *If you do not already have an online identity, you are offered the option to create one via HMRC using documents that you might have at home, such as your passport and P60 (a record of a person's income and tax for the previous year). This means that you may not have to go to visit the Job Centre in person. To use this option, you need to input your passport number and the amount of money you were paid in the previous tax year into a secure online system. HMRC will then do an automated check with the Passport Office and its own systems, and let DWP know if they are able to confirm your identity.*
- *If you do not have a passport or P60, you can choose to input information from other financial documents such as bank statements into HMRC's secure system in order to create an online identity. HMRC will then do a one-off, secure identity check with a financial agency such as TransUnion, as these hold records for most people.*
- *The project described here is intended to make processes easy and usable by more people than existing government identity checking systems which require people to have a passport, P60 or to have registered with HMRC.*
- *Some groups say that the identity checks described here exclude some people, such as those without credit histories, with unusual residences or residence histories, or otherwise complicated lives. This means that such people might be under more scrutiny and this might delay their access to Universal Credit.*

After these statements, respondents were informed that the DWP is exploring possible extra checks that it could use in the future to make access to verified online accounts more secure, and they were asked to record their views about these possible extra checks on the same slider used above ranging from ‘Very comfortable’ to ‘Not at all comfortable’. The checks listed were:

- *whether the time you log in is similar to your usual log in time*
- *whether the rhythm that you type your password is the same as your usual rhythm*
- *whether you swiped on your phone in the same pattern as you normally swipe it*
- *whether you are using the same device that you have used for previous interactions with the system.*

The distribution of responses to these items is shown in Figure 21. The curves here can be read in the same way as those in Figure 17, as the sliders were presented in the same way, with differences in the specific statements to which they were asked to respond.

Figure 21: Responses to statements about DWP Confirm Your Identity

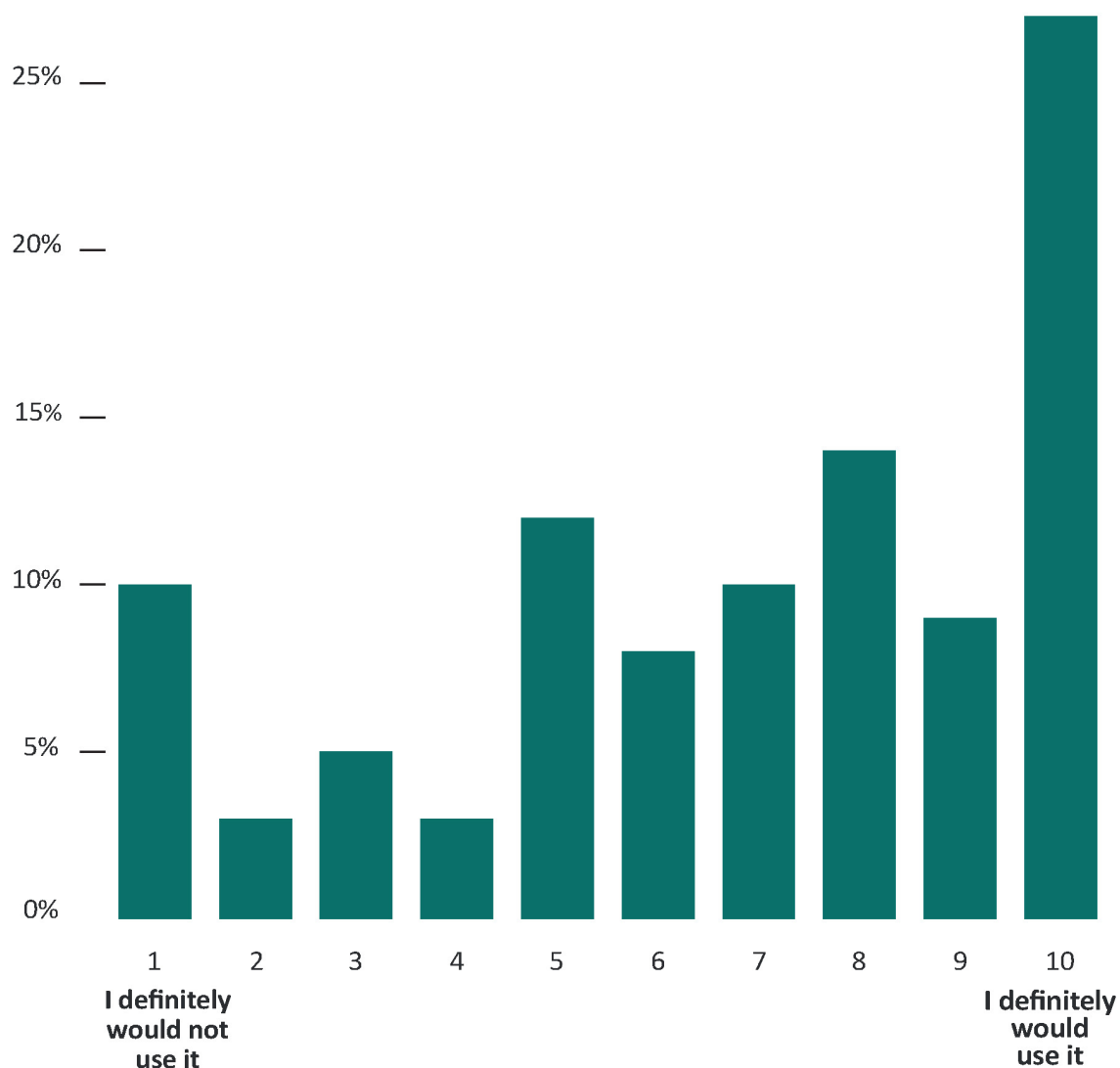


Respondents were largely comfortable with **DWP Confirm Your Identity**. Many respondents accepted the purpose of this data practice, and in free text fields, around 1/4 respondents indicated that they would use Confirm Your Identity because it is convenient. But some aspects of DWP Confirm Your Identity concerned respondents, such as those that might reinforce or lead to inequalities. They were least comfortable about the statement that there is concern that some groups of people might be excluded, put under more scrutiny, or have their access to Universal Credit delayed by the introduction of online identity verification. Respondents were also not so comfortable in response to the question about possible extra security checks. The check with which they were least comfortable was checking whether people logging into accounts swipe their phones in the same pattern as they normally swipe it.

While most groups had similar attitudes towards these questions, whether people were claiming Universal Credit made a difference. People claiming Universal Credit felt significantly more comfortable about additional automated checks than other respondents. Although they were not as positive about potential additional checks as they were about the other statements, people who are not in receipt of Universal Credit were even less positive.

We asked respondents how likely they would be to use the identity verification process we described, if they had to claim Universal Credit at the DWP. The results are shown in Figure 22. Just over a quarter of respondents reported that they definitely would use an online ID rather than taking their documents into the Job Centre to be verified. The free text fields suggest this is down to convenience, with around a quarter of comments noting that the online verification system would be easier and more convenient to use than the current system. Only 10% of respondents stated that they definitely would not use it.

Figure 22: If you had to claim Universal Credit at the DWP, how likely would you be to use an online ID rather than taking your documents into the Job Centre to be verified?



The only difference in people's likelihood of using Confirm Your Identity was by income, with people who were more highly-paid broadly more likely to state that they would use it if they had to claim Universal Credit. However, it is worth noting that this group may be the least likely of all of the income groups to *have to* use it.

Finally, as with other sections of the survey about specific data uses, respondents were asked how surprised they were to learn about identity verification at the DWP, and how much they felt they understood it. Responses to these questions are shown in Figures 23 and 24. Roughly equal fractions of respondents stated they were either surprised or not surprised by the details of this data practice: 48% were very surprised (15%) or fairly surprised (33%), and 53% were not surprised (39% not very, 14% not at all). As with other specific data practices, a significant majority of respondents stated they somewhat understood the programme.

Figure 23: Are you surprised at the details of how the DWP is using people's data?

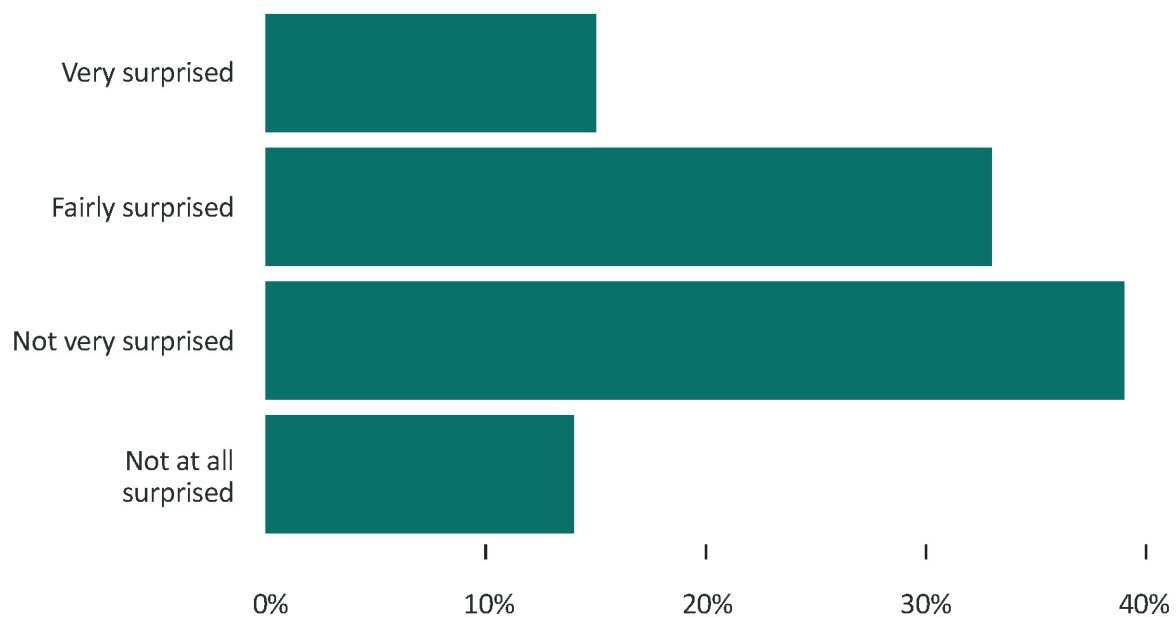
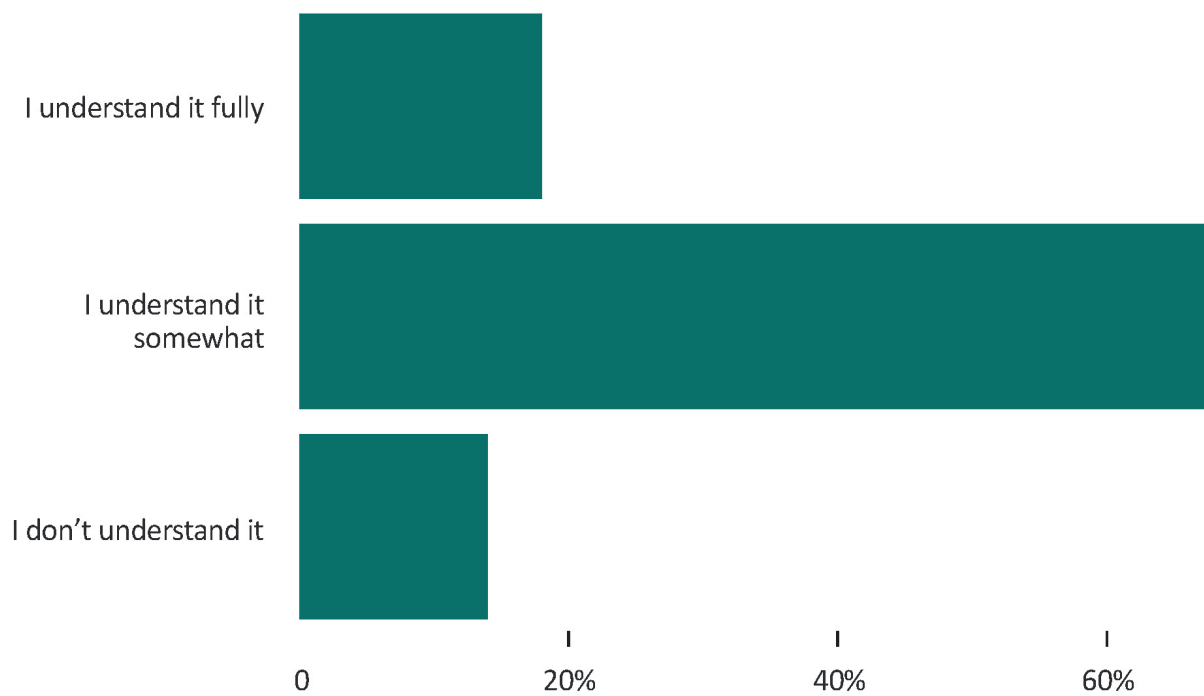


Figure 24: How much do you feel you understand how the DWP is using people's data?



Women were more surprised to find out about Confirm Your Identity than men, as were younger people compared to older people. The differences in surprise by qualifications were less pronounced, but people with intermediate qualifications – vocational qualifications, and those qualifications taken at 16 or 18 – were less likely to be surprised, while those with either no qualifications or higher educational qualifications were more likely to be surprised.

As with the other examples of public sector data uses, there were differences in understanding by qualifications, income, and age, consistent with the other examples. There were also differences by gender, but not by ethnic group (unlike the NHS Covid data store, where there were differences by ethnic group). Younger people (23% compared to 12% of older people), men (22% compared to 15% of women), people in higher income brackets (32% compared to 14% of people in lower income brackets), and people with higher levels of education (23% compared to 11% of people with lower levels of education) were more likely to state that they understand Confirm Your Identity. The differences by education were less pronounced than for the other data practices.

10. Classifying awareness and attitudes

Are people who are more knowledgeable more or less concerned about data uses? Given widespread claims about the importance of more transparency about data uses (for example, by the Office for National Statistics, and the Information Commissioner's Office in the UK), the answer to this question matters. To explore the relationship between (i) people's awareness of data practices and (ii) their attitudes towards them, we produced classifications of awareness and attitudes. We used latent class analysis to do this, based on our analysis of responses, rather than how respondents described themselves.

We found four clusters of awareness, which we call:

- *Knowledgeables* (who generally correctly identify true and false statements);
- *Believers* (who generally respond that the statements they are presented with are true);
- *Disbelievers* (who generally respond that statements are false, even when they are in fact true);
- *Don't knows* (who generally state that they don't know the answer to awareness questions).

We found four clusters of attitudes, which we call:

- *Critical* (who strongly disagree with some statements that are positive about data practices, and strongly agree with others that are negative about data practices);
- *Cautious* (who tend to agree or disagree in the same directions as the *Critical* group, but not strongly);
- *Neutral* (who overwhelmingly respond with 'Neither agree nor disagree');
- *Agree* (who overwhelmingly either agree or strongly agree with all statements).

We arrived at these findings by generating two classifications, based on responses to the awareness and attitude questions that we asked in the first part of the survey and that are discussed in sections 5 and 6 of this report. In each case, we used latent class analysis to classify respondents into a number of different latent classes, using the **poLCA** package in R. Latent class analysis is a subset of structural equation modelling, used to estimate latent groups from a number of categorical variables. Models were estimated for between 2 and 10 latent classes in each case, with each number estimated ten times, and with 100,000 iterations each time.

There are various approaches to model selection in latent class analysis, with a range of statistical criteria, as well as theoretically-driven approaches. In both cases, having inspected different classifications, we opted for four-class models. In the absence of a strong model-based solution, we opt both for models that are straightforward to interpret and with classes with a reasonably large fraction of the sample, in order to be able to investigate differences. Further details of our process can be found in Appendix 13.2 below.

10.1. Latent class model: knowledge/awareness

Figures 25-26 shows the distribution of responses within each of the knowledge variables for the latent classes based on these variables. The upper panel addresses the variables around questions about ways in which data is collected and used, while the lower panel addresses the variables around true/false questions, all of which are discussed in (5) above.

Figure 25: Latent class model (awareness): first set of variables relating to awareness questions about data collection and use



Figure 26: Latent class model (awareness): second set of variables relating to true/false questions



We call the four awareness clusters *Knowledgeables*, *Believers*, *Disbelievers*, and *Don't knows*. The largest cluster at 38%, *Believers* are more likely than any other group to state that the false statements in the second batch of questions are in fact true.

The next largest cluster at 25%, *Knowledgeables* mostly got the questions about data collection and use right, and they also generally correctly identified the false statements in the second batch. Again, this was not uniformly the case. For example, the majority believe that phones are used to track people's eye movements to track what they look at online.

The smallest cluster, at 14%, *Disbelievers* are the opposite of the *Believers*. They generally responded that statements are false, even when they are in fact true. Again this is not uniformly the case. For example, the majority correctly recognized that data is used to identify what people like them like to do online, but the overall pattern was to define statements as false.

Finally, *Don't knows*, at 23%, generally stated that they didn't know the answer to questions, rather than providing a definitive answer. Once again, this is not the case across the board, with some statements receiving more definitive answers than others, but this group is distinguished by its large number of don't know responses.

10.2. Latent class model: attitudes

We call the four attitude clusters *Critical*, *Cautious*, *Neutral* and *Agree*. The two largest groups were *Critical* (36%) and *Cautious* (34%). These groups are not distinguished by which questions they agree or disagree with most. Rather, they are distinguished by their strength of feeling. Large fractions of the *Critical* group strongly disagree with some items, and strongly agree with others. By contrast, the *Cautious* group tends to agree or disagree, but not strongly.

The *Neutral* group (13%) overwhelmingly responds with 'Neither agree nor disagree'. Again, this varies by question. Notably, a moderate fraction of them reports that they don't have strong opinions about the collection and use of data about them, but in general, they hold neutral positions.

Finally, the *Agree* group (17%) overwhelmingly either agrees or strongly agrees with all questions. This seems contradictory – for example, a respondent simultaneously strongly agreeing that they don't have strong opinions about the collection and use of data about them, and that they're concerned about the role of commercial companies in public services – but it may be that some people within this group find the items difficult to interpret, or use the "Strongly agree" category to communicate their strength of feeling more generally.

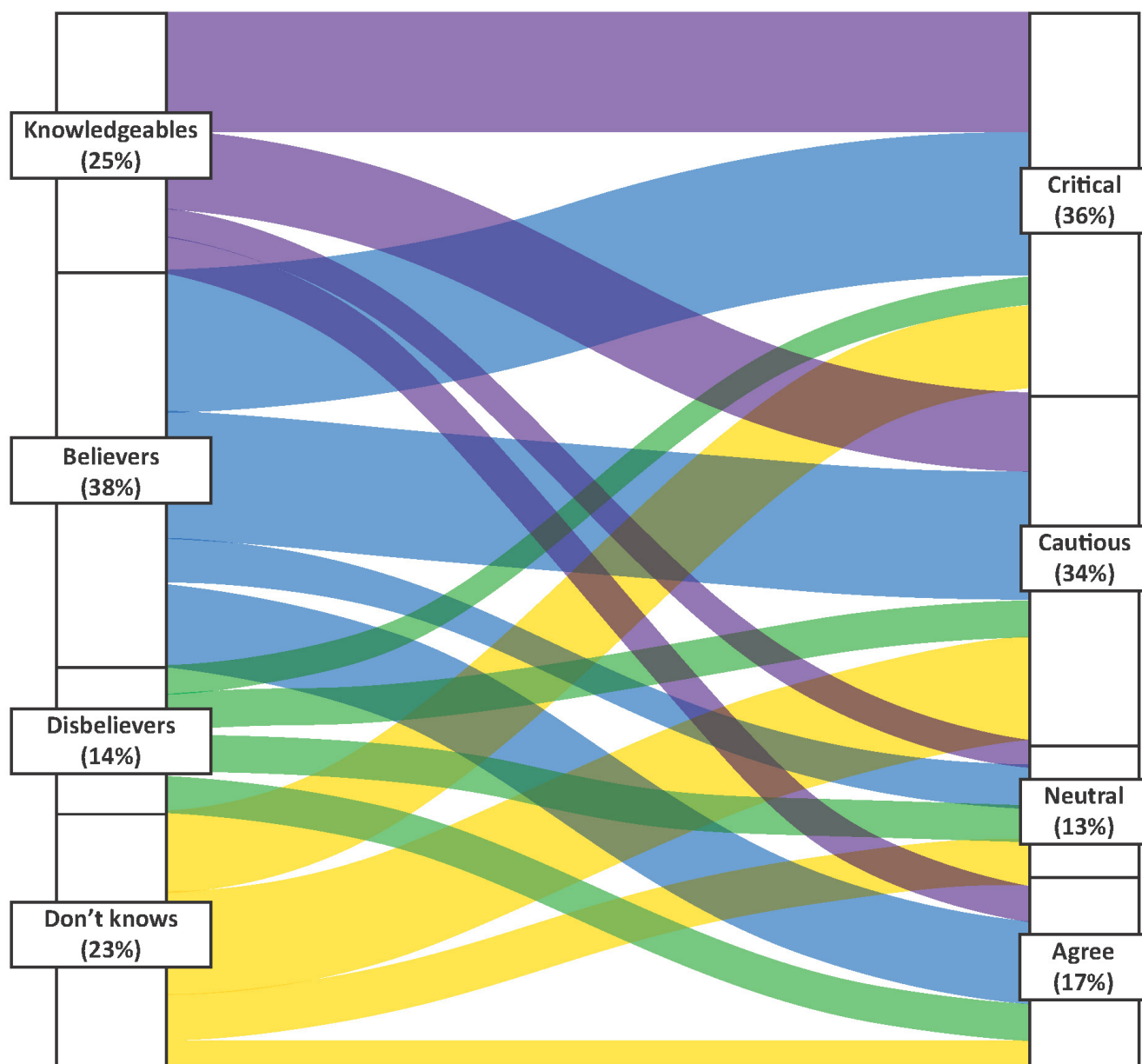
Figure 27: Latent class model (attitudes)



10.3. The relationship between knowledge and attitude classifications

Figure 28 illustrates the relationships between the two models. It shows that *Knowledgeables* are more likely to be in the *Critical* group, and moderately more likely to be in the *Cautious* group. There is also a clear relationship between respondents who are *Neutral* in attitude (they generally neither agree nor disagree) and *Don't knows* in awareness (who generally stated that they didn't know the answer to questions).

Figure 28: The relationship between knowledge clusters (on the left) and attitude clusters (on the right). The bars connecting knowledge clusters to attitude clusters show the overlap between them.



10.4. Awareness and attitude classifications in relation to public sector data practices

Awareness and attitude classifications relate to responses to questions about specific public sector data practices in these ways:

- Respondents' likelihood of using **BBC Box** varies by both classifications, awareness and attitude. There are clear differences between groups in both sets of clusters. *Knowledgeables* are by far the least likely to use BBC Box, while the other groups' responses are largely similar to each other. This pattern is echoed in the attitude clusters, with *Critical* respondents most likely to say they definitely would not use it, although with smaller differences than by awareness. Respondents who were *Cautious* and *Neutral* in attitudes have relatively large fractions who neither definitely would nor definitely would not use BBC Box, while the *Agree* group is the most positive about using BBC Box of any cluster.
- The differences in attitudes towards the specific dimensions of **BBC Own It** amongst the awareness clusters are relatively small. *Knowledgeables* are most positive about the fact that anonymous data collection can be turned off, and in general the most likely to report that they were not at all comfortable with specific dimensions of Own It. Other groups' responses are broadly similar to each other. The differences between attitudes clusters are more pronounced, with the *Critical* cluster consistently least comfortable with all dimensions.
- In relation to preferred options for possible future sharing of **BBC Own It** data with university researchers, differences by awareness cluster are relatively small, save for the fact that *Disbelievers* (who are most likely to respond that particular data uses do not happen) are less likely to have "No sharing" as their preferred option. By contrast, there are larger differences by attitude cluster. *Critical* and *Neutral* clusters are far more likely to prefer the "No sharing" option. The *Agree* group (which overwhelmingly either agrees or strongly agrees with statements) is by far most in favour of other possible future data sharing options.
- Members of different knowledge clusters largely had similar responses to statements about **DWP Confirm Your Identity**, save for statements about automated checks, where *Knowledgeables* were far less comfortable than the other groups. The *Critical* attitude group was less comfortable across the board. In some cases, these differences were less pronounced, but for statements about automated checks, these differences were large. The *Agree* attitude group was the most comfortable with all statements.
- The differences in how likely respondents were to use **DWP Confirm Your Identity** by knowledge cluster were relatively small, whereas differences by attitude cluster were more pronounced. The *Agree* group were by far the most positive attitude group of all. The *Critical* group were the least likely to use Confirm Your Identity.
- In relation to **surprise at** the specific public sector data uses by the different classifications, in the awareness clusters, *Knowledgeables* are consistently the least surprised, followed by *Believers*. The size of these differences varies, and is largest for BBC data uses, compared to NHS and DWP data uses. *Don't knows* are less likely to be surprised than other groups about the details of the NHS Covid data store. By attitude cluster, the *Neutral* group is least likely to be surprised of all groups, and the *Agree* group is most likely to be surprised.
- In relation to **comfort with** the specific public sector data uses by the different classifications, in the awareness cluster, *Disbelievers* are the most comfortable, while the *Believers* are only slightly behind them for the NHS Covid data store. *Knowledgeables* are the least comfortable with data uses. By attitude cluster, the *Agree* group are by far the most comfortable on both measures, while the *Cautious* and *Neutral* groups are fairly similar. The critical group are by far the least comfortable with data uses.

- In relation to **understanding of** the specific public sector data uses by the different classifications, in the awareness cluster, patterns are inconsistent. *Knowledgeables* are most likely to report that they understand the BBC data uses, but second least likely to report that they understand the NHS Covid data store and DWP Confirm Your Identity, behind the *Don't know* group. *Don't knows* are consistently the group most likely to report that they don't understand the specific data uses presented to them in the survey. In the attitude cluster, respondents in the *Neutral* group are most likely to report that they don't understand each set of data uses, while the *Agree* group are by far the most likely to report that they understand data uses fully.

10.5. Demographic difference and classifications

We explored the relationships between cluster membership and demographic factors using **multinomial logistic regression** in order to separate out the direct effects of each demographic variable, rather than produce comparisons via cross tabulation. For example, it might be that the most important thing that predicts whether someone is a member of the *Critical* group is whether they are highly-educated. However, because people in older age cohorts are less likely to have degrees than younger people (ONS 2019), the differences by age may appear exaggerated.

We used cluster membership as our dependent variable, and age, disability, qualifications, ethnic group, and gender as our independent variables. For the model predicting knowledge, we use *Knowledgeable* as the reference category, while for the model predicting attitudes, we used *Critical* as the reference category. Results are presented graphically, with point estimates as points and 95% confidence intervals surrounding them. We report multinomial logistic coefficients, rather than odds ratios, in each case.

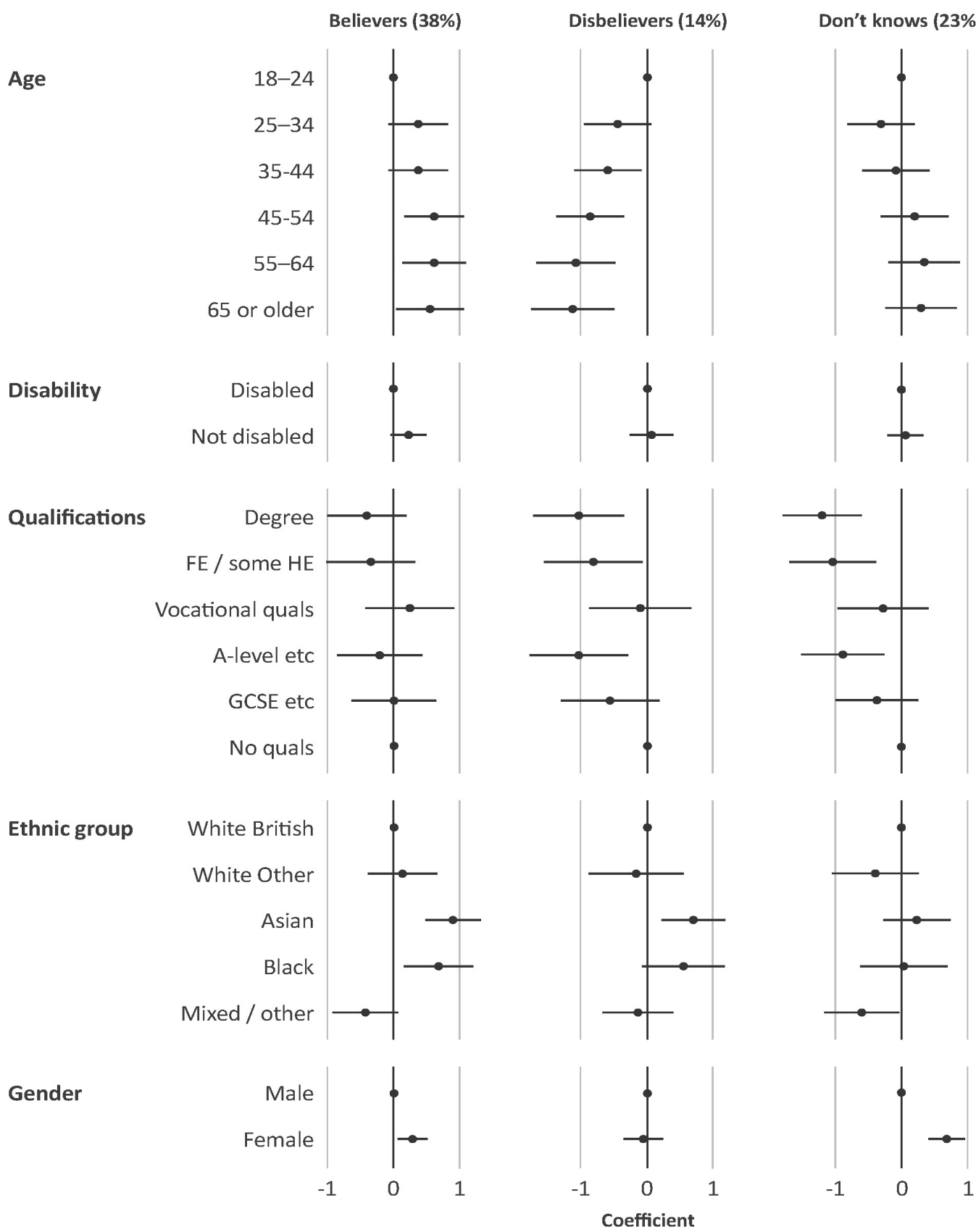
Modelling awareness

Figure 29 shows that there are differences in predicted awareness cluster membership by demographic group. Younger people are least likely to be in the *Believers* cluster (who generally respond that the statements they are presented with are true). People aged 45 or older are significantly more likely to be *Believers*. By contrast, younger people are most likely to be *Disbelievers*, while members of each older group are progressively less likely. Older people are slightly more likely to be *Knowledgeables*.

Black and Asian people are more likely to be in the *Believers* group. They are also more likely to be in the *Disbelievers* group, but this difference is only significant for Asian people. White people and people of Mixed/other ethnicities are more likely to be *Knowledgeables*. Women are marginally more likely to be in *Believers* than men, and much more likely to be in *Don't knows*, while men are more likely to be *Knowledgeables*.

There are no differences between disabled people and people who aren't disabled in the awareness clusters. There are no significant differences across educational qualifications amongst *Believers*, but those with higher education or qualifications taken around the age of 18 are less likely to be *Disbelievers* or *Don't knows* than people with vocational qualifications, qualifications taken around the age of 16, or no qualifications. More highly qualified groups are more likely to be *Knowledgeables*.

Figure 29: Multinomial logistic regression predicting knowledge cluster
(base outcome: knowledgeable)

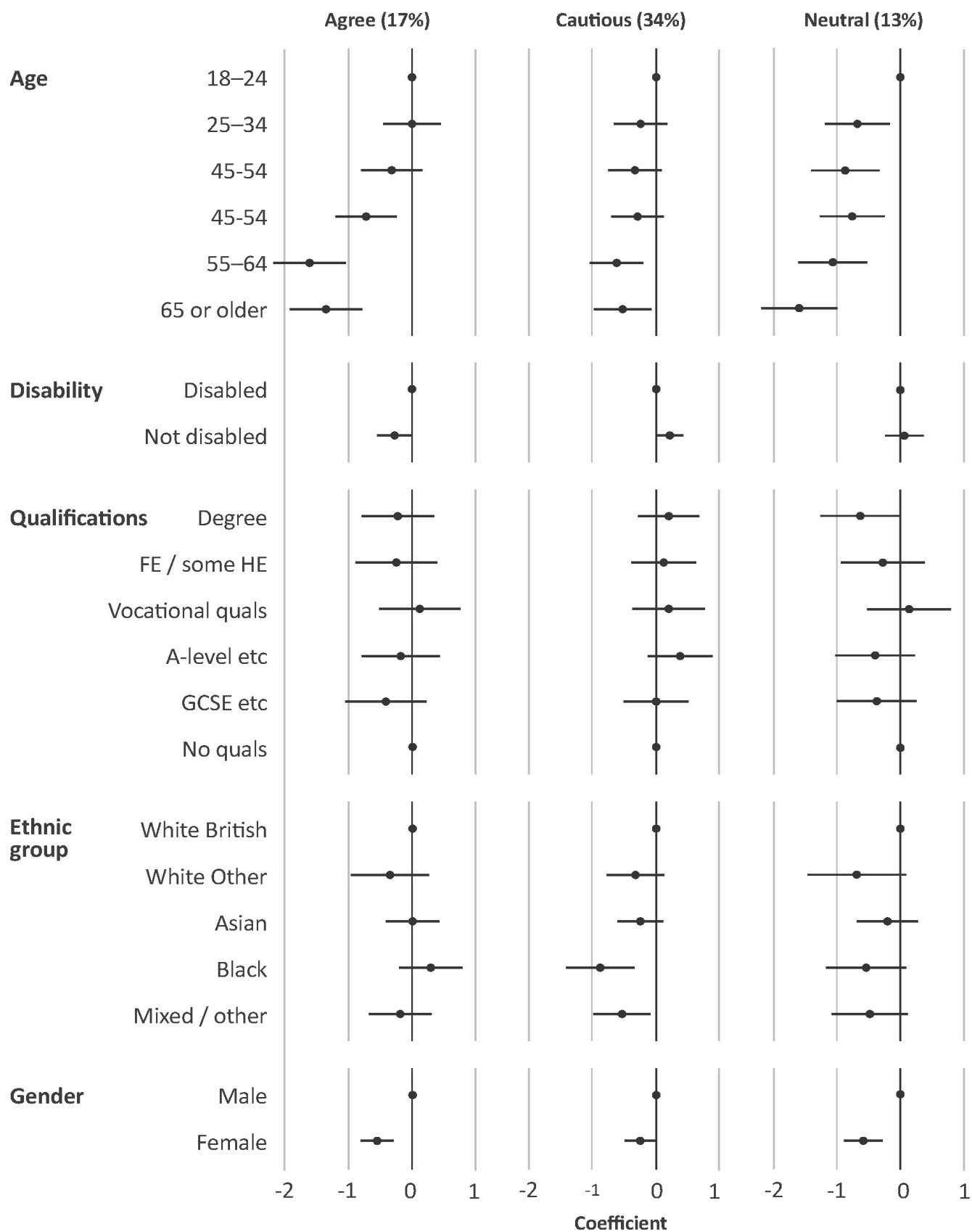


Modelling attitudes

Figure 30 shows that there are differences in predicted attitude cluster membership by demographic group. Younger people are much more likely than older people to be in the *Agree* cluster – this is the largest difference of all in relation to the attitude classification. This is most striking for those aged 55-64, who are least likely to be in this cluster. A similar pattern can be seen for the *Neutral* cluster, though with smaller differences. Younger people are marginally less likely to be in the *Cautious* cluster and older people are much more likely to be in the *Critical* cluster. Women are much less likely to be in the *Agree* and *Neutral* clusters than men, and slightly less likely to be in the *Cautious* cluster and so are more likely to be in the *Critical* cluster.

As with knowledge, there are no differences between disabled and non-disabled people in terms of attitudes. There are also no differences between different qualifications in terms of attitudes. This is in contrast to knowledge, where differences were larger. There are no differences by ethnicity, except that Black people and people of Mixed/other ethnicities are less likely to be in the *Cautious* cluster.

Figure 30: Multinomial logistic regression predicting attitude cluster
(base outcome: critical)



11. CONCLUSIONS & RECOMMENDATIONS

Public concern about certain data uses remains high. Survey respondents indicate high levels of concern with particular aspects of data use and data uses in particular contexts. For example, across different parts of the survey, we found that people do not like commercial companies profiting from personal data, especially if this data was originally gathered for pro-social purposes such as public health. Even where we found statistical support for data uses like the NHS Covid data store, in written comments respondents expressed significant concern – in this case, about commercial companies profiting from or misusing data. Similar high levels of concern have also been identified in previous research. This consistent finding communicates a strong message to data policy-makers and practitioners about public dissatisfaction with existing data uses.

It is important to note that in our survey, we sought to explore public attitudes to data uses, whether positive or negative. We worded our questions carefully and made sure to include a balance of statements about the potential benefits as well as the potential harms of data uses. We did not look for concerns alone. Given this, it is noteworthy that we find high levels of concern about data uses across all demographic groups: it is not the case that only groups that may be disadvantaged or harmed by certain data uses are uncomfortable with them.

Context is a defining factor when it comes to public attitudes to data uses. Some data uses are more concerning than others. Commercial companies profiting from personal data, organisations tracking when, where and how people log on, and automated technologies being used to try to change people's behaviours are more concerning than personal data being used to manage Covid-19. It is therefore not helpful to talk about public attitudes to data uses out of context, or to assume that public views of data uses will be consistent across contexts. Rather, we need to be attentive to contextual specificity. Furthermore, complexities arise when less concerning data uses come together with those that are more concerning. In the case of the NHS Covid data store, respondents support its pro-social, public health aims and are simultaneously concerned about the involvement of commercial companies and the lack of clarity about the ways in which these companies may have access to or profit from personal data in the future.

Within different contexts, there are differences in attitudes depending on the type of data use in question. Respondents were broadly comfortable with BBC Own It, but concerned about some details. Likewise, they were broadly comfortable with DWP Confirm Your Identity, but they had some concerns about some of the possible extra checks that DWP was exploring, such as checking whether people logging into accounts swipe their phones in the same pattern as they normally swipe it, even though these are designed to enhance security.

It is also important to note that context influences trust. Therefore, trust issues go beyond the specific data use. People trust the health care sector a lot more than they trust technology, media and social media companies, regardless of the type of data use. So high trust in GP and NHS data uses is in danger of being diminished by the involvement of less trusted parties, such as technology companies, in health data initiatives like the NHS Covid data store. Generally, degree of trust in data use results from attitudes to organisations or sectors more generally, so to address *data* trust issues, change needs to go beyond data uses and to focus on trustworthiness more generally.

People want to know more about data uses, but the more they know, the more they are concerned about many of them. People who are more knowledgeable about data uses are more likely to have negative attitudes towards them. How we interpret this finding is important. We should not conclude that we therefore shouldn't tell people about data uses! People want information about what happens to their personal data: they want to know who has access to data about them and where data about them is stored. In the absence of clear information about data uses, people imagine negative scenarios in which their data may be sold, leaked or misused, and this leads to concern and distrust. Put together, these findings tell us that change is needed.

There are some **differences in attitudes to data uses across demographic groups**. Disabled people support health data uses more than people without disabilities, and people claiming Universal Credit felt more comfortable with data-driven identity checks than non-claimants, for example. The biggest differences are between older and younger people. Younger people are more confident online, more likely to be online almost constantly, and more likely to do a wide range of things online. They are more likely to state that they understand the specific different data practices and are slightly more positive about them. Older people were more likely to express surprise at BBC data uses, while younger people were more likely to express surprise at DWP Confirm Your Identity, and the different age groups had similar responses to the NHS Covid data store. Older people were more likely to be in the Critical cluster in relation to their attitudes to data uses, while younger people were more likely to be in the Neutral cluster. In the awareness cluster, younger people were more likely to be Believers, while older people were more likely to be Disbelievers. But none of these differences is enormous. In terms of people's attitudes towards data uses, **different demographic groups have a lot in common**. It isn't the case that the groups that are most affected by data practices are the only ones concerned about them

Future data policy-making and data practice should be informed by our findings, for example:

- **Consider NOT engaging in practices** that consistently concern the public, for example sharing data originally intended for health or other prosocial purposes in ways which enable commercial companies to profit.
- **Communicate clearly and fully** about existing data uses and possible future uses, especially where such uses go beyond what was originally intended, because in the absence of clear information, people imagine negative and concerning future data use scenarios. Doing this requires investment of time and effort.
- **Change practices**. The people who know most about data uses are the most concerned about them. If data uses continue unchanged, the public will continue to be concerned, regardless of clarity of communication.
- **Consult the public**, particularly those affected by data uses, about what changes to data uses are needed. Engaging in a genuine dialogue with the public about data uses on a case-by-case basis, because public views of data uses differ based on context.
- Understand that **some trust issues go beyond the type of data use** and instead result from attitudes to organisations or sectors more generally. To address *data* trust issues, fundamental changes may be needed that go beyond data uses and practices.

12. REFERENCES

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

- doteveryone. (2018a). People, power and technology: The 2018 digital attitudes report.
<https://understanding.doteveryone.org.uk>.
- doteveryone. (2018b). *People, power, and technology: The 2018 digital understanding report*.
<https://understanding.doteveryone.org.uk>
- Draper, N and Turow, J (2019) The corporate cultivation of digital resignation. *New Media & Society*.
- Eubanks, V. (2017). *Automating inequality: How high-tech tools profile, police and punish the poor*. St Martins Press.
- ICO (Information Commissioner's Office) (2018) Information Rights Strategic Plan: Trust and Confidence.
<https://ico.org.uk/media/about-the-ico/documents/2259732/annual-track-2018.pdf>
- Kennedy, H, Hartman, T, Steedman, R and Jones, R (2020) UK public unhappy with the way their personal data is managed. <https://livingwithdata.org/project/wp-content/uploads/2020/03/Views-on-Data-Management-Full-Report.pdf>
- Kitchin, R (2014) *The Data Revolution: big data, open data, data infrastructures and their consequences*, Sage Publications.
- ONS (Office for National Statistics) (2019) Highest qualifications held in age groups, by economic activity for UK, 2018.
<https://www.ons.gov.uk/employmentandlabourmarket/peopleinwork/employmentandemployeetypes/adhocs/10635highestqualificationsheldinagegroupsbyeconomicactivityforuk2018>

13. APPENDICES

13.1. The survey

A full list of our original survey questions can be found on our Living With Data website, at this link:
<https://livingwithdata.org/resources/living-with-data-survey-results/>.

13.2. Classification modelling

There are various approaches to model selection in latent class analysis, with a range of statistical criteria, as well as theoretically-driven approaches. To illustrate how the different models vary, figure 1 shows an alluvial diagram of how cases are allocated to different classes as the number of classes increases, in this case drawn from the attitudes classification. Cases are coloured according to their class membership at ten classes. For example, we can see that, as the number of classes increases from three to four, almost all the observations in the class at the top in the four-class model are derived from the class at the top of the three-class mode, that almost all of the observations in the second class in the four-class model are derived from the observations in the second class in the three-class model, and that almost all of the observations in the bottom class in the four-class model are derived from the bottom cluster in the three-class model. We can further see that the observations in the third class in the four-class model are derived roughly equally from the middle and bottom classes in the three-class model.

Figure 1: indicative alluvial diagram

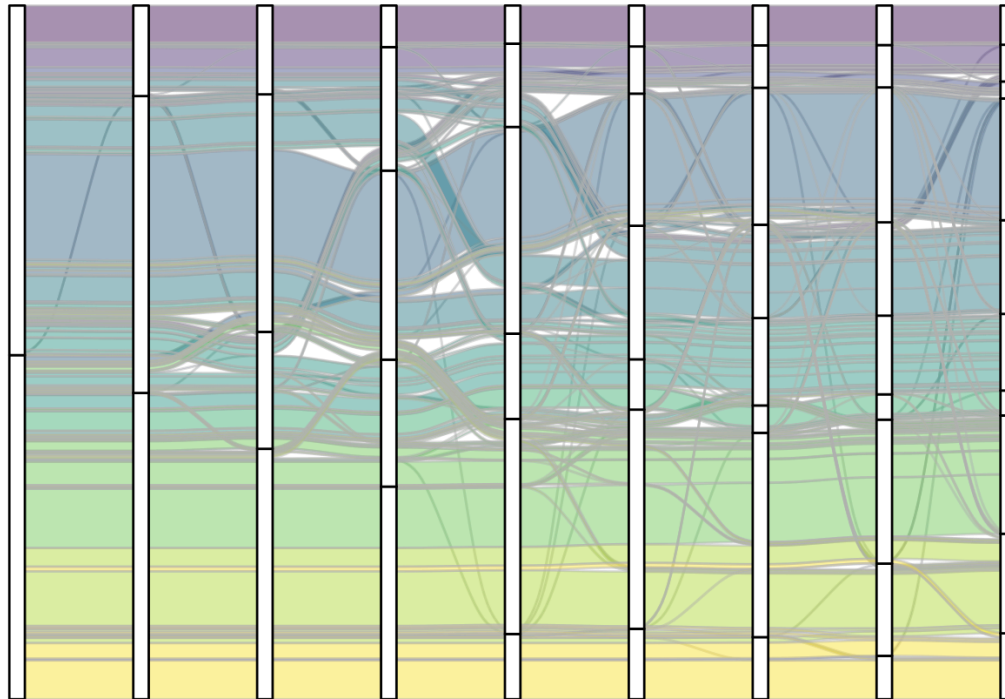
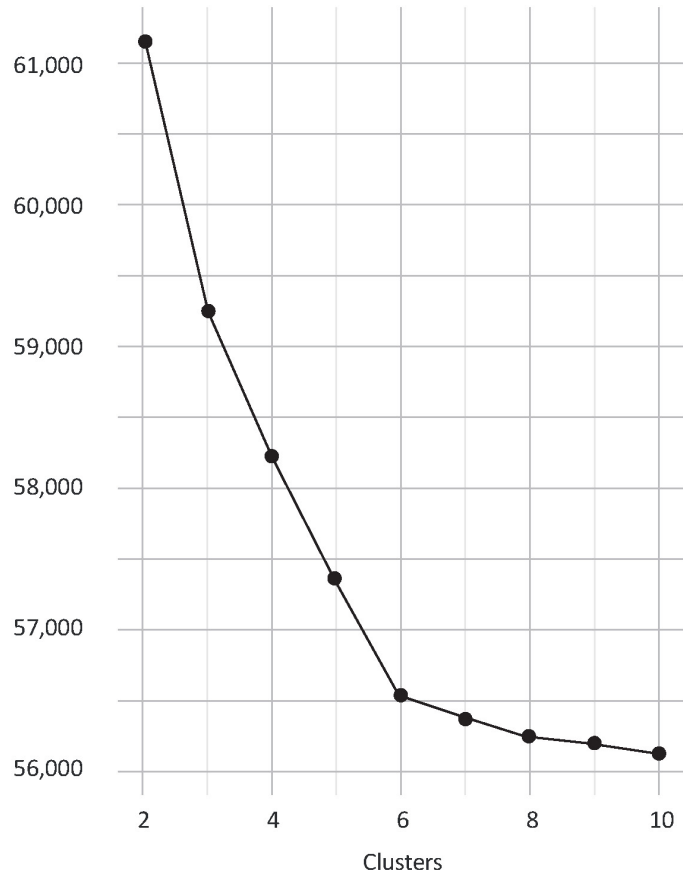


Figure 2 shows the Bayesian Information Criterion for both sets of models. The models of attitudes reach a local minimum at seven clusters, while the models of knowledge decrease as the number of clusters increases, with a “knee” at six clusters. However, in both cases the range of BIC values is relatively small.

In both cases, having inspected the different classifications, we opt for four-class models. In the absence of a strong model-based solution, we opt both for models that are relatively straightforwardly interpretable, and with classes with a reasonably large fraction of the sample, so as to be able to investigate differences.

Figure 2: BIC scores for the latent class models

Knowledge



Attitudes

