



This is a repository copy of *Challenging assumptions about the relationship between awareness of and attitudes to data uses amongst the UK public.*

White Rose Research Online URL for this paper:

<https://eprints.whiterose.ac.uk/203288/>

Version: Published Version

---

**Article:**

Taylor, M. [orcid.org/0000-0001-5943-9796](https://orcid.org/0000-0001-5943-9796), Kennedy, H. and Oman, S. (2024) Challenging assumptions about the relationship between awareness of and attitudes to data uses amongst the UK public. *The Information Society*, 40 (1). pp. 32-53. ISSN 0197-2243

<https://doi.org/10.1080/01972243.2023.2283729>

---

**Reuse**

This article is distributed under the terms of the Creative Commons Attribution (CC BY) licence. This licence allows you to distribute, remix, tweak, and build upon the work, even commercially, as long as you credit the authors for the original work. More information and the full terms of the licence here:

<https://creativecommons.org/licenses/>

**Takedown**

If you consider content in White Rose Research Online to be in breach of UK law, please notify us by emailing [eprints@whiterose.ac.uk](mailto:eprints@whiterose.ac.uk) including the URL of the record and the reason for the withdrawal request.



[eprints@whiterose.ac.uk](mailto:eprints@whiterose.ac.uk)  
<https://eprints.whiterose.ac.uk/>

# Challenging assumptions about the relationship between awareness of and attitudes to data uses amongst the UK public

Mark Taylor, Helen Kennedy & Susan Oman

**To cite this article:** Mark Taylor, Helen Kennedy & Susan Oman (29 Nov 2023): Challenging assumptions about the relationship between awareness of and attitudes to data uses amongst the UK public, The Information Society, DOI: [10.1080/01972243.2023.2283729](https://doi.org/10.1080/01972243.2023.2283729)

**To link to this article:** <https://doi.org/10.1080/01972243.2023.2283729>



© 2023 The Author(s). Published with license by Taylor & Francis Group, LLC.



Published online: 29 Nov 2023.



Submit your article to this journal [↗](#)



View related articles [↗](#)



View Crossmark data [↗](#)

# Challenging assumptions about the relationship between awareness of and attitudes to data uses amongst the UK public

Mark Taylor<sup>a</sup>, Helen Kennedy<sup>b</sup> and Susan Oman<sup>c</sup>

<sup>a</sup>Sheffield Methods Institute, University of Sheffield, Sheffield, UK; <sup>b</sup>Department of Sociological Studies, University of Sheffield, Sheffield, UK; <sup>c</sup>Information School, University of Sheffield, Sheffield, UK

## ABSTRACT

This article advances understanding of the relationship between (a) people's awareness of and (b) their attitudes toward the ways in which data about them is collected, analyzed, shared, and used. It draws on an online survey of 2000 adults in the UK, which found that people with greater awareness of data uses hold more negative attitudes toward them. This finding is important because it challenges the deficit model which underlies initiatives that seek to improve the public's attitudes toward and trust in institutional data uses through improved transparency or better data literacy.

## ARTICLE HISTORY

Received 28 October 2022

Accepted 10 November 2023

## KEYWORDS

Awareness; data policy; data uses; public attitudes; survey methods

## Introduction

Personal data “related to an identified or identifiable person” (European Union 2016) is increasingly ubiquitous, commercially valuable, and central to decision-making. Policy initiatives across the globe seek to ensure that such data is collected, analyzed, and shared in ways that are ethical and responsible and which minimize harms, e.g. in the UK's National Data Strategy (Department for Digital, Culture, Media and Sport 2020). One way to do this, it is argued (e.g. Hartman et al. 2020), is to factor in public views about “data uses,” a shorthand term we use to refer to institutional data collection, analysis, and sharing. As a result, studies of people's awareness of and attitudes toward data uses have proliferated in recent years.

Despite the increase in research into awareness of and attitudes to data uses, the relationship between the two things, awareness and attitudes, is rarely considered. This relationship is important because it is sometimes assumed that the more people understand data uses, the more positive their attitudes will be (CDEI 2021). This assumption underlies advocacy and policy initiatives across various domains of the data society. For example, data literacy projects often conceive of the “data trust deficit” (RSS 2014) as a problem that results from a deficit of understanding. Elsewhere, transparency advocates argue that more

transparent information about how data-driven systems is needed, to enable understanding and democratic scrutiny of them (Pasquale 2015).

The assumption that better understanding will lead to more positive attitudes mobilizes a “deficit model” (Irwin 2014), or a belief that the public has gaps in its knowledge which, if filled, will lead to more positive attitudes. The deficit model has been criticized by scholars from various disciplines. Science communication and public understanding of science scholars argue that its focus on improving public trust, rather than improving system trustworthiness, is misplaced (e.g. Aitken, Cunningham-Burley, and Pagliari 2016). Critical race scholars posit that it fails to acknowledge the role that structural inequalities play in shaping public attitudes (e.g. Benjamin 2016). In this article, we apply these critiques to the emerging field of data studies, advancing understanding of the relationship between people's awareness of and attitudes to data uses through original empirical research.

A further problem with research into public “awareness of” and “attitudes to” data uses relates to the terms used, hence our use of quotation marks around them here. Understanding, knowledge, and awareness are sometimes used as if they are synonymous. The same is true of attitudes, opinions, and perceptions. But despite their interchangeable use in everyday life, the meanings of terms in each cluster are not

identical. This matters because of the conclusions that are drawn from research undertaken and the policy decisions that follow. Recognizing this as a problem means it is necessary to ask questions, such as do accurate responses to true/false statements about data uses tell us about understanding, knowledge, awareness, how well-informed people are, or something else? Methodological choices about what to ask research respondents have consequences for what we can accurately claim to have found. Reflecting on the precise meanings of these sets of terms is therefore important. In what follows, we use awareness and attitudes as shorthand for the clusters of terms listed above, whilst simultaneously interrogating their appropriateness for describing the research that we undertook.

In this article, we report findings from an online survey of 2000 adults in the UK, undertaken at the end of 2020, which offers a response to our research question: what characterizes the relationship between people's awareness of and attitudes toward data uses? In the survey, we collected primary data about different aspects of people's awareness of and attitudes toward data uses, with a particular focus on public sector data uses in the contexts of health, welfare, and public service broadcasting. We used latent class analysis to explore patterns of responses and analyze the relationship between awareness and attitudes. We found that people with greater awareness of data uses hold more negative attitudes toward them. This finding is important because it challenges the "deficit model" (Irwin 2014) assumption that greater understanding of data uses will result in more positive attitudes toward them.

We proceed by situating our research in the context of broader debate about what the public know, think, and feel about data uses and reflection on the varied terms used in research in this area. We then discuss our methods and findings. We conclude by reflecting on the implications of our findings, for knowledge, policy, and practice.

### **Prior research into people's awareness of and attitudes to data uses**

There is a wealth of literature, academic and grey, on public views of data uses, which emerges from a wide range of disciplines and countries (see Kennedy et al. 2020 review for an extended summary). Studies of public views of data uses tends to focus *either* on attitudes *or* on awareness. Research into attitudes identifies widespread public concern about what happens to personal data. In 2019, the UK Information

Commissioner's Office's annual report claimed that "a record number of people" were "raising data protection concerns" (ICO 2019). Concerns relate to a range of issues, including how companies use people's personal data (in the US, Auxier et al. 2019); who is mining personal data (in Australia, Lupton and Michael 2017); commercial organizations controlling people's personal data in return for the digital services they provide (in the UK, Hartman et al. 2020). Research consistently finds that context is an important factor: people are less concerned about healthcare institutions using their personal data (in Ireland, Robinson and Dolk 2015), and much more concerned about marketers, advertisers, and social media companies doing so (in the UK, ICO 2019). Concern varies depending on the kinds of data and platforms in question (in the US, Rendina and Mustanski 2018) and across different demographic groups (in Sweden, Bergstrom 2015; in the UK, Edwards, Gillies, and Gorin 2021). Despite this variation, conclusions point in the same direction: people are concerned. And yet people continue to engage with data-driven systems, in what Draper and Turow (2019) describe as "digital resignation," or reluctant acceptance of data uses (Peppin 2020).

A largely separate set of studies focuses on understanding or awareness. Findings from this work tend to be mixed: people are well-informed about certain data uses and less informed about others, as elaborated in the paragraph below. This is not surprising, given the variety of data uses that exist. Understanding and awareness of data collection, analysis, sharing, and re-use—the various processes we bring together under the shorthand term "data uses"—can include: understanding regulation, or what companies and governments are legally permitted to do and legally forbidden from doing (e.g. European Commission 2019); definitions, or what different terms mean (Woodruff et al. 2018); practices, or what different organizations actually do with data (Doteveryone 2018a, 2018b; Eslami et al. 2015); and awareness of the societal benefits of data uses (Ditchfield et al. 2002). Research into people's awareness of data uses addresses all of these issues and finds that awareness varies significantly across domains and practices.

The introduction of the EU's General Data Protection Regulation (GDPR) in 2016 generated a flurry of research into people's understanding of their rights in relation to data uses, as enshrined in the new legislation. Results are conflicting: one survey found understanding of data rights was relatively low in the UK (ICO 2019), whilst another found it was relatively high compared with other countries in

Europe (European Commission 2019). Other research has examined what people think companies do, or can do, with personal data. Doteveryone (2018a), a think tank focusing on responsible technology, found that most people were aware that data is collected about their online searches, the websites they visit and their online purchasing history (68, 68, and 70% of their survey respondents, respectively), but far fewer people knew that data about their internet connection and information other people share about them are also collected (38 and 17%, respectively). Findings from qualitative research confirm this mixed picture. One study of Facebook users found they had limited awareness of the platform's data uses (Eslami et al. 2015), whereas another argues that people's behaviors suggest they are aware (Bucher 2017).

There are few exceptions to this pattern of researching *either* attitudes *or* understanding. Doteveryone has researched both but reports separately on each issue (2018a, 2018b). Kennedy, Steedman, and Jones's (2020) qualitative research into attitudes to data gathering is another exception. They found that older participants and younger participants with mild learning disabilities did not fully understand the data uses discussed with them, but they nonetheless had strong, largely negative attitudes to them. They argue that full understanding is not necessarily a prerequisite to developing views on data uses.

### Assumed deficits

Despite the general absence of research that explicitly explores the relationship between awareness of and attitudes toward data uses, assumptions about the relationship exist, for example in data transparency and data literacy scholarship and advocacy. Within data transparency scholarship, it has been proposed that more transparent information about how data-driven systems work will enable democratic scrutiny of them, and individuals can then make informed decisions about their participation in said systems (Pasquale 2015). Alongside fairness and accountability, transparency is widely seen as a foundational principle for responsible data science (Facctconference.org 2018). However, assumptions about the normative value of transparency are increasingly called into question, for example in Ananny and Crawford's widely cited paper on the limitations of the transparency ideal (2018). Critics argue that making data-driven systems transparent is extremely difficult in practice, that transparency can be used for public relations purposes rather than to enhance understanding, and that arguing for transparency can imply an acceptance

of high-risk data systems which should be questioned (see Bates et al. 2023 for a summary). Nonetheless, transparency remains a central tenet in responsible data policy and practice (CDEI 2021).

Data literacy initiatives, or "educational interventions to improve citizens' data literacies" (Yates et al. n.d.), like transparency endeavors, also assume that the "data trust deficit" (RSS 2014) is a problem that results from a deficit of understanding. Examples include The Data Literacy Project (thedataliteracy-project.org) and School of Data (schoolofdata.org), both of which conceive of skills for making sense of data as empowering. Where transparency addresses the data trust deficit with better information, data literacy addresses it by equipping individuals with better skills to make sense of available information. Like data transparency endeavors, data literacy projects are also subject to criticism, for their narrow focus on technical or numeric rather than critical capabilities (Gray, Gerlitz, and Bounegru 2018), prioritizing of individual over collective needs (Sander 2021), or failure to conceive of literacy as the ability to understand the world in order to change it, as popular educator Freire (1996) did D'Ignazio (2017).

Data transparency and data literacy initiatives both mobilize a "deficit model" (Irwin 2014), in their assumption that filling gaps in public knowledge will lead to positive outcomes, such as more informed choices or greater empowerment. In scholarship on public understanding of science, such models have been criticized for their attention to improving public attitudes, rather than improving the objects to which said attitudes are directed (Aitken, Cunningham-Burley, and Pagliari 2016). Criticisms of the deficit model can also be found in relation to education (Gorski 2011) and cultural participation (Miles and Sullivan 2012), and yet it remains foundational across much policy and practice.

The work of critical race scholars, such as Ruha Benjamin, also challenges the assumption that more understanding of systems results in more positive attitudes toward them. She uses the term "informed refusal" to describe what she witnessed in her research into Black people's engagement in health projects (2016). Benjamin argues that the notion of informed consent wrongly assumes that "the transmission of information" results in "the granting of permission" (2016, 967). In racialized communities, this assumption does not always hold up, and what we see instead is informed refusal. Thus we need to unpack "the racial logics of trust" (2014, 755), writes Benjamin.

Benjamin's argument advances critiques of the deficit model by applying an anti-racist lens and

acknowledging the role that structural inequalities play in shaping public attitudes. It suggests there is no inevitable relationship between understanding and trust, acceptance, or other positive attitudes. Our research also advances these criticisms, as we discuss below.

### Troubling the terms used in research on data use “attitudes” and “awareness”

A further problem with research into public attitudes to data uses relates to terms used and claims made. As noted above, understanding, knowledge, and awareness are often used interchangeably, and so are attitudes, opinions, and perceptions. All of these terms have multiple meanings, and while the terms in each cluster may overlap, they are not synonymous. Yet public attitudes research rarely acknowledges this or reflects on precisely what can be claimed about people’s responses to data uses based on the research that has been undertaken. For example, none of the literatures cited thus far defines the terminology that it uses. Doteveryone (2018a) presents a model of digital understanding in their report on their survey, which classifies respondents as aware (the lowest level of understanding), discovering (the middle level), or questioning (the highest level). They describe understanding as moving from “basic awareness” to “deeper questioning” and thus see awareness as limited understanding, but they do not define any of these terms. And yet, reflecting on the multiple and overlapping meanings of all terms—understanding, knowledge, awareness, attitudes, opinions, and perceptions—is an important pre-requisite to making claims about what people know, understand, feel, or opine about data uses. We do this below, to frame our own claim-making about our survey findings.

Oman’s (2021) work is an exception to the general pattern of not reflecting on or defining terms. She presents numerous meanings of the term “understanding” and applies them to the context of data uses. The first conception of understanding relates to how individuals comprehend data and the understanding they have of situations involving data uses. Such understanding is not universal. For some people, data uses are complex and opaque; others have an abstract, intellectual understanding of them; and others still, an everyday comprehension of their application. The second conception of understanding is more collective; it relates to our shared understanding, highlighting the difficulties of ensuring we are talking about the same thing to others with different understandings. The third conception relates to empathy, where data

uses are comprehended *with understanding* for others, or without it.

Drawing on empirical research into citizens’ data literacies, Yates et al. (n.d.) turn to the other two terms in the awareness cluster, knowledge, and awareness, and differentiate them as follows:

Knowledge of the details of both overt and covert data collection, sharing and trading by platforms and other organizations may be quite limited for many users, this does not mean they are not aware it is happening. In all the focus groups respondents expressed an awareness that data around use of platforms is collected; but what, how, why and which organizations are involved with it – were often poorly understood (p. xi).

In contrast to sociology of knowledge approaches which often perceive knowledge as sets of ideas accepted by a group or society (Doyle McCarthy 1996), Yates et al. adopt a more everyday understanding of knowledge, perceiving it as based on facts or information. As used here, awareness overlaps with knowledge, but it is different from knowledge. It is more akin to being conscious of something than to deep understanding or knowledge of facts. Whereas awareness is aligned with feelings or perceptions, knowledge is more aligned with facts and information. Knowledge not only implies depth of understanding of external events or information, but also a sense that what is understood is “true” or “fact” (Oman 2022). In the context of data uses, arriving at knowledge can be hard, because they are opaque, dynamic, subject to frequent change, and speculation. Thus awareness as generalized knowledge, rather than full understanding of details, might be easier to achieve than knowledge. So awareness is different from knowledge and understanding, yet these differences are rarely acknowledged in literature on public attitudes to data uses.

Attitudes, opinions, and perceptions are more or less dependent on understanding, knowledge, or awareness. Oskamp and Schultz (2014) claim that attitudes are the ways that people perceive the world around them. Myers defines an attitude as “a favorable or unfavorable evaluative reaction toward someone or something” (2012, 36; see also Bem 1970). For Oskamp and Schultz, attitudes emphasize evaluation and learning, aid decision-making, influence thinking, and action, and explain consistency in people’s behaviors. They are simultaneously comprehensive and simple. Opinions are understood as views or judgments formed about something, not necessarily based on fact or knowledge. Knowledge is therefore not a pre-requisite to attitudes or opinions. Indeed, researchers from fields, such as the sociology of emotions



have long highlighted the important role that feelings play in the formation of attitudes and opinions and their epistemological value (e.g. Ahmed 2004). Likewise, in psychology and the cognitive sciences, perception is understood as the process of getting, interpreting, selecting, and organizing *sensory* rather than *cognitive* information (IOMC n.d.).

Being reflective about these various, sometimes overlapping definitions of terms facilitates precision in the conclusions that are drawn from research into public views of data uses and the policy and practice decisions that follow. Being reflective involves asking questions like if people are asked whether specific statements about data uses are true or false and they answer accurately, does that tell us about their knowledge, awareness, how well-informed they are, how much they understand, or something else? If people are asked how strongly they agree or disagree with statements about data uses, are we finding out about attitudes, opinions, perceptions, feelings? Do the answers to these two questions vary, depending on the precise wording of the statements?

In the analysis of our survey data that follows, we explore the nature of the relationship between awareness and attitudes, treating neither as monolithic, given that research has shown that diverse publics are better-informed about some data-related issues than others. Throughout our analysis, we reflect on whether our findings are telling us about understanding, knowledge, or awareness, and about perceptions, attitudes, or opinions. We have used the terms awareness and attitudes thus far in the article and in its title, as these are widely used and serve as a shorthand for the two clusters of terms. In what follows, we consider whether they are the most suitable terms. In so doing, we contribute a reflective lens where such terms have largely been used unquestioningly. Finally, we attend to demographic differences, acknowledging the role that structural inequalities play in awareness of and attitudes toward data uses. We start by introducing our data and methods.

## Data and methods

Data was collected in September and October 2020, via the web survey platform Qualtrics, which recruited participants as well as hosting the survey. The sample was recruited to be nationally representative of adults in the UK, in relation to gender, age, income, and disability. There was additional recruitment (or “boosts”) of people born outside the UK, LGBTQ+ people, Black, Asian, and other racialized people, and people in receipt of the UK’s main welfare benefit,

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 people with low educational qualifications, due to the complexity of qualifications in the UK. As a consequence of this, and because people who choose to opt into unfiltered, web-based surveys tend to hold more qualifications (Gelman et al. 2016), our respondents have slightly higher qualifications than is nationally representative. Table 1 provides demographic information about respondents.

We applied two filters to address the data quality issues that often emerge in online surveys in which respondents are paid for their participation. The first was by using free text fields. Participants were asked to share their additional thoughts about questions they had been asked at two points during the survey. Respondents whose answers suggested they were not paying attention, either by entering unintelligible (e.g. “tfjyhgfdh fkjm m”) or irrelevant content (e.g. “Marijuana cigarettes weed marijuana”), were removed from the sample. The second was a speed check. After a pilot launch with  $N=40$  to confirm question intelligibility, respondents who completed the survey in less than half the median time taken by pilot respondents—that is, less than 450 s—were removed. After these data quality checks, the overall sample size was 2000.

Respondents were asked to read an information sheet, confirm they had read it, and that they agreed to participate in the research, before undertaking the survey. Ethical review, through the University of Sheffield Research Ethics Committee, via the Department of Sociological Studies (application number 032273) was approved on January 6, 2020.

Our analysis in this article focuses on two sets of questions within the survey questionnaire. The first set relates to awareness. In this section, respondents were presented with statements about different ways in which organizations collect data; different ways in which organizations use data; and general, factual statements, including definitional ones. Respondents were asked to state whether statements were true, false, or don’t know. The second set of questions related to attitudes toward data uses. As with the awareness questions, these covered the range of processes we include in our term “data uses”: that is, collection, analysis, sharing, and uses of data. Respondents were asked to indicate how much they agreed or disagreed with each statement. We included a set of demographic questions in the survey, to explore differences between demographic groups, and

**Table 1.** Demographic information about respondents.

| Variable                        | Category                                           | Proportion (%) |
|---------------------------------|----------------------------------------------------|----------------|
| Highest education qualification | None/don't know                                    | 6              |
|                                 | Post-16 qualifications (e.g. GCSEs)                | 27             |
|                                 | Post-18 qualifications (e.g. A-levels)             | 19             |
|                                 | Some higher education (e.g. 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 in household           | Any                                                | 36             |
| Country of birth                | UK                                                 | 87             |
|                                 | Other                                              | 13             |
| Citizenship                     | UK                                                 | 93             |
|                                 | Other European country                             | 4              |
|                                 | Other non-European country                         | 4              |
|                                 | Reported disability                                | 27             |
| Disability                      |                                                    |                |
| Ethnic group                    | White British                                      | 72             |
|                                 | White other                                        | 5              |
|                                 | Black                                              | 6              |
|                                 | Asian                                              | 11             |
|                                 | Mixed ethnicities/other                            | 6              |
|                                 | Male                                               | 48             |
|                                 | Female                                             | 51             |
| Sexuality                       | Non-binary                                         | 1              |
|                                 | Heterosexual/straight                              | 80             |
|                                 | LGBTQ+                                             | 20             |
| Employment/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              |
|                                 | £19,999 or less                                    | 29             |
| Total household income          | £20,000–£39,999                                    | 38             |
|                                 | £40,000–£59,999                                    | 21             |
|                                 | £60,000–£100,000                                   | 10             |
|                                 | £100,000 or more                                   | 2              |

we refer to answers to these questions in our analysis below.

Writing questions about data uses across a wide range of domains so that they are understandable by a general audience is challenging. In line with standard practice in survey design, we built on questions from other polls in this section of our survey. However, agreeing the precise wording of these questions was not straightforward. For example, some of our questions about whether organizations collect or

use data about people in particular ways originated in Doteveryone's (2018b) digital attitudes survey. But whereas Doteveryone's wording was "Do you think that organizations collect/use data about people in the following ways?" we used "any organizations," because without "any," respondents may interpret the question as referring to all or most organizations. Although this would make comparing our results with Doteveryone's difficult, we felt that this was important to clarify our meaning in this way.



Devising plausible false statements about data collection and data use that are not far-fetched is particularly challenging, but it is also necessary to avoid presenting respondents with a long list of only true statements. One reason it is challenging is because it is hard to be certain that statements are in fact false. In our team, we disputed whether some statements, identified as either true or false in other surveys, were in fact so. For example, we were not all convinced that the correct answer to “do any organizations collect data by tracking people’s eye movements to track what they look at online” was “no,” the answer deemed to be correct in the survey which had used this question (Doteveryone 2018b). Innovations like eye-tracking—which have certainly been trialed even if they are not widely implemented—make it hard to know unequivocally whether such statements are true or false, especially when the question is about *any* organizations. Despite our uncertainty, we included this question with “no” as the correct answer, following Doteveryone. We reflect on the implications of this decision below.

Below, we first present descriptive statistics of answers to the questions in the awareness and attitudes batteries, to understand awareness and attitudes broadly. Second, we use latent class analysis, a statistical procedure that we used to identify different subgroups within each set of answers, and to investigate the relationship between our classifications of awareness and attitudes. Third, we estimate whether certain groups are more or less likely to be members of some latent classes or others. We do this to understand the role of demographic differences and social inequalities in awareness of and attitudes toward data uses.

For our latent class analysis, we used the poLCA package in R. 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. Having inspected different classifications, we opted for four-class models for both our awareness and our attitudes questions. In the absence of a strong model-based solution, we selected models that were both straightforward to interpret and which had classes with a reasonably large fraction of the sample, to be able to investigate differences. Figure A1 (Appendix A) reports figures for a range of goodness-of-fit statistics, illustrating the lack of a clear model-based solution for either set of questions. The scores cannot be meaningfully compared across the models due to the different numbers of variables in

each case. However, the shapes of the distributions are very similar across all four criteria.

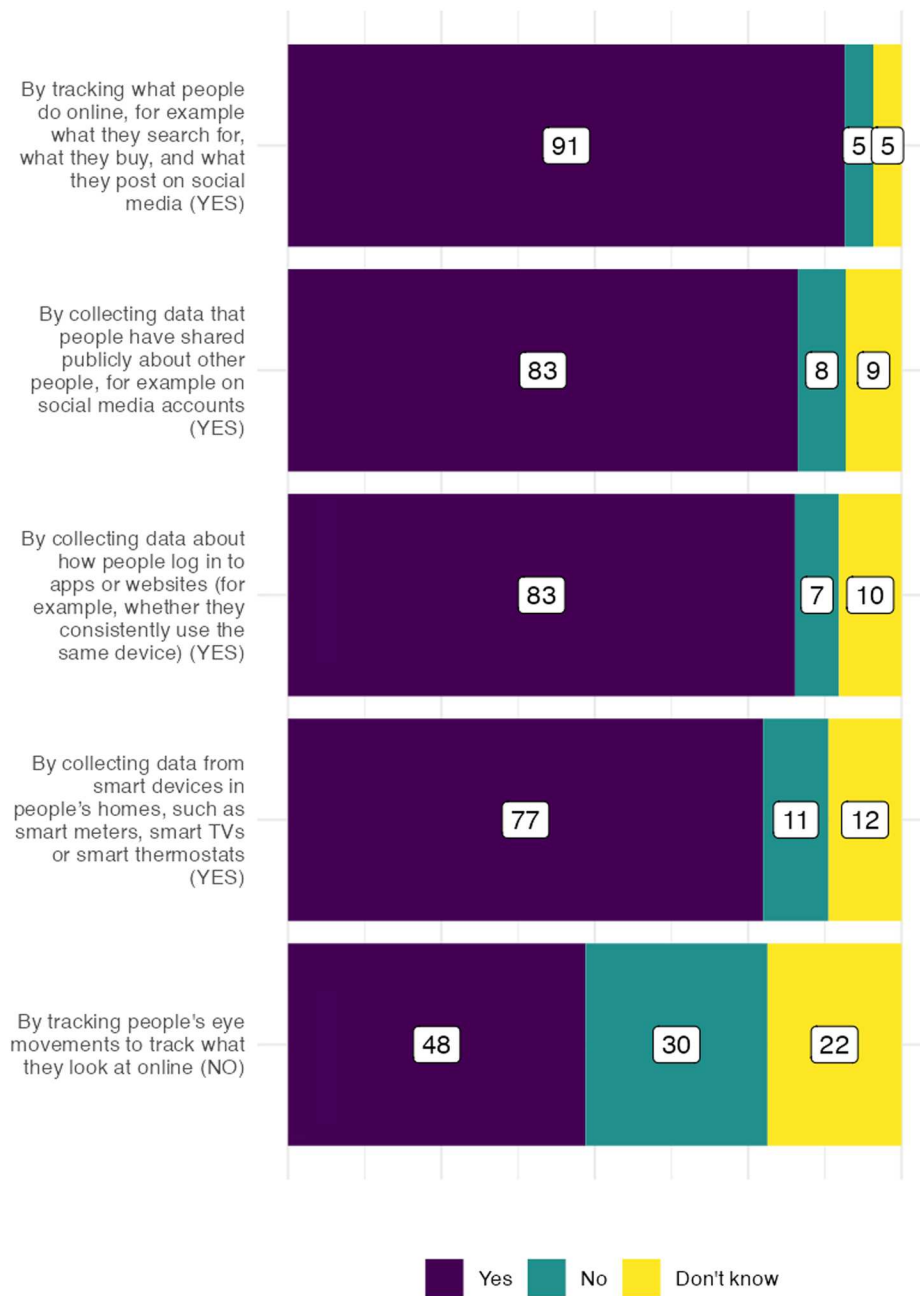
## Results

### *Descriptive statistics: Awareness of and attitudes to data uses*

We presented respondents with five statements relating to the question “Do you think that any organizations *collect* data about people in the following ways?” and five statements for “Do you think that any organizations *use* data about people in the following ways?” We asked respondents to indicate whether they thought that the statements were true, false, or that they didn’t know. We then estimated their awareness of data practices based on their answers, as other surveys of public awareness of data uses do (e.g. Auxier et al. 2019 for Pew; Doteveryone 2018a, 2018b; European Commission 2019; ICO 2019). Statements, which were presented in randomized order, and responses can be seen in Figures 1 and 2.

A majority of respondents correctly identified the true statements about data collection as such, while 48% believed the false statement (“organizations collect data by tracking people’s eye movements to track what they look at online”) was true. However, as noted above, we were also not convinced that this was not true, so respondents’ uncertainty about this mirrored our own. In most cases, there were similar numbers of people responding that they didn’t know and that statements were false. Our addition of the word “any,” described above, to the question “Do you think that organizations collect/use data about people in the following ways?” may account for the slightly higher numbers of positive responses to statements that both we and Doteveryone used, such as “By tracking what people do online” (91% in our survey, 85% in Doteveryone’s), “By collecting data that people have shared collectively” (83%/57%) and “By collecting data from smart devices” (77%/60%).

In the next question about data uses, all five statements were identified as true by Doteveryone (2018a) where they originated. The majority of respondents correctly identified them as such. Larger fractions of respondents stated that they thought that three data uses either didn’t occur or that they didn’t know whether they occur; namely, that data is used to help the government keep people safe; to suggest that people do things differently to improve their well-being; to help protect people from scams. Results are shown in Figure 2.



**Figure 1.** Do you think that any organizations collect data about people in the following ways?

We then asked respondents to indicate whether a series of statements were true, false, or that they didn't know. Of the nine statements presented, four were true and five were false. The statements and results are shown in Figure 3. Here we can see that of the five false statements, only one, "banks sometimes send their customers emails asking them to click links to verify their accounts," was correctly identified as such by a majority of respondents (66%). A majority of respondents (56%) 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. The three other false statements had similar numbers of people responding true, false, and don't know. Respondents were more likely to state that false statements were true than vice versa: three out of four true statements were correctly identified as such by a majority of respondents, compared to only one out of five false statements. On average, 26% of people responded "don't know" to false statements, compared with 23% for true statements. We suggest this indicates more of a reluctance to state categorically that data uses *do not* occur than that they *do* occur.

We told respondents the correct answers to awareness questions before they moved onto the attitudes section. To estimate attitudes, we presented respondents with ten statements and asked them to indicate whether they strongly agreed, agreed, neither agreed nor disagreed, somewhat disagreed, or strongly disagreed with the statements. These statements were prefaced with a note that encouraged respondents to answer honestly and that stated that there are no right or wrong answers to these questions. Statements and responses are presented in Figure 4.

Respondents wanted to know who has access to data about them (83% agreed or strongly agreed with the relevant statement) and where data about them is stored (80%), and they wanted more control over how their data is used by organizations (83%). They did not support corporate profit-making from personal data (60% disagreed or strongly disagreed with the relevant statement), and only 26% of respondents stated that they don't have "strong opinions about the collection and use of data about me." Together, we argue that these responses indicate high levels of concern about data uses, confirming the findings of other studies (e.g. Hartman et al. 2020).

At the same time, 52% of respondents agreed or strongly agreed that collecting and analyzing data can be good for society. This shows that whilst they have some concerns about data uses, people also recognize the potential benefits. 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 can also be seen as an indication of hesitation or concern.

### Classifying awareness and attitudes

It is possible that the people who answered questions about one issue correctly, and thus indicated a degree of awareness about that issue, may be less aware about other issues and therefore answer related questions incorrectly. It is also possible that people with concerns about a particular data-related issue are not equally concerned about all issues. Some people might be very concerned about certain issues, whereas others may be concerned about a broad range of issues. We generated two classifications using latent class, based on responses to the awareness and attitude questions, respectively, and in each case, we classified respondents into four different latent classes.

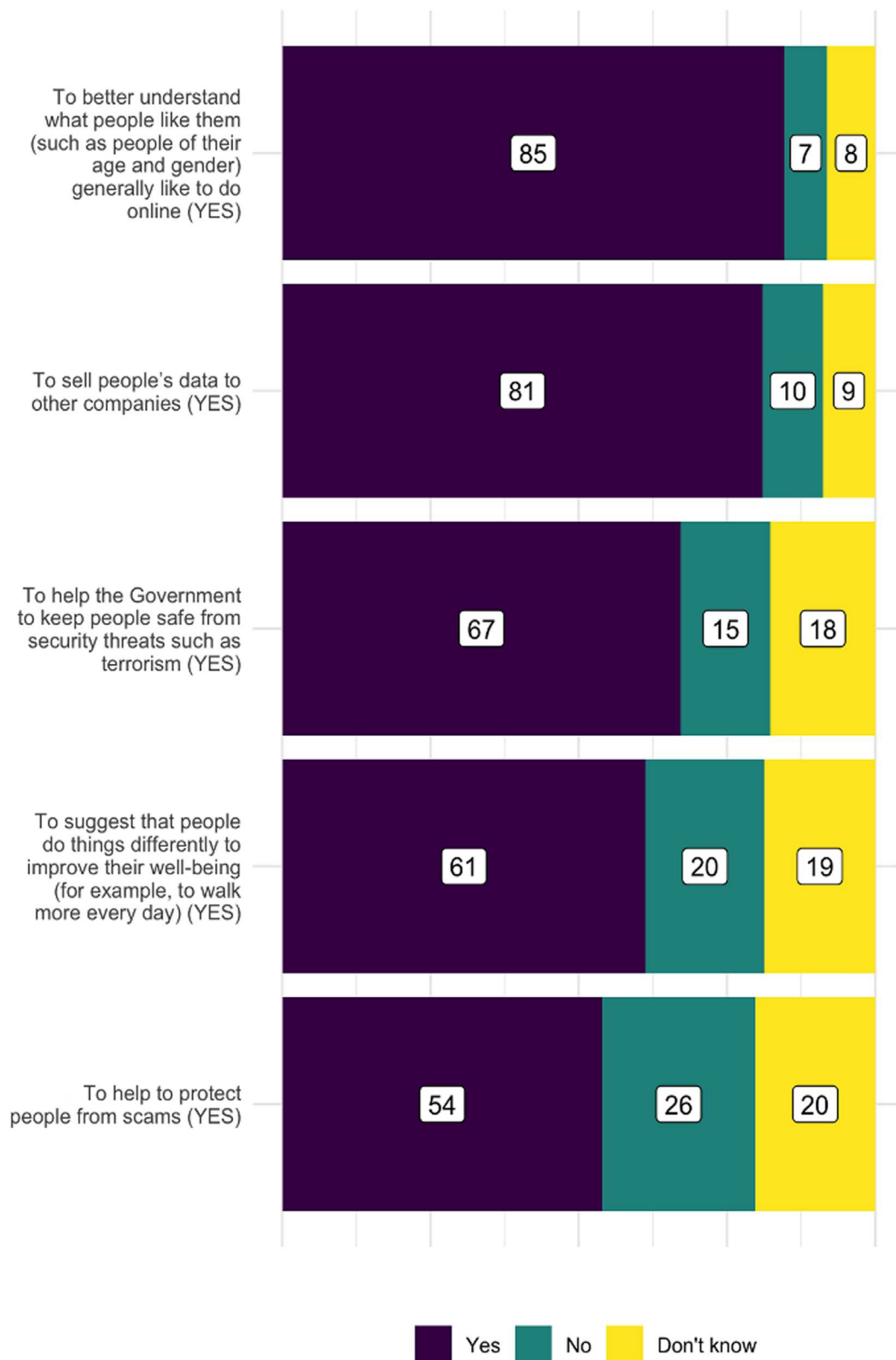
Based on awareness questions, we describe our four latent classes as *Aware*; *Believers*; *Disbelievers*; *Don't knows*. Based on attitudes questions, we describe

our four latent classes as *Critical*; *Cautious*; *Neutral*; *Agree*. All classes are described in Table 2 below, where percentages of respondents in each class are also listed.

Figures 5 and 6 show the differences between the four latent classes based on the awareness variables, by illustrating the distribution of responses within each latent class.

The largest latent class, *Believers*, at 38%, are more likely than any other group to state that false statements are in fact true. The next largest latent class, *Aware* (25%), mostly answered the questions about data collection and use correctly. They generally correctly identified the false statements in the true/false batch, with the exception of the statement about which we were also uncertain—namely, that organizations track people's eye movements to track what they look at online. The smallest latent class, *Disbelievers* (14%), are the opposite of *Believers*. They generally responded that statements are false, even when they were in fact true. Again, this was not uniformly the case. For example, the majority correctly stated 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* (23%) generally stated that they didn't know the answer to questions, rather than providing a definitive answer. Some statements received more definitive answers than others, but this group is distinguished by its large number of don't know responses.

Figure 7 shows the results of the latent class based on attitudes, showing the distributions of responses to each attitude question within each of the four latent classes. The two largest groups are *Critical* (36%) and *Cautious* (34%). These groups are not distinguished by the questions with which they agree or disagree. Rather, they are distinguished by their strength of feeling. Large fractions of the *Critical* group strongly disagreed with statements, such as "I support corporate profit-making from personal data" and "I don't have strong opinions about the collection and use of data about me." They strongly agreed with statements that indicated that they wanted more control over how their personal data is used and to know who has access to data about them and where it is stored. By contrast, the *Cautious* group tended to hold similar opinions to the *Critical* group, but not strongly. The *Neutral* group (13%) overwhelmingly responded with "Neither agree nor disagree." Again, this varied by question. Notably, a moderate fraction of the *Neutral* group agreed with the statement that they don't have strong opinions about the collection and use of data about them.



**Figure 2.** Do you think that any organizations use data about people in the following ways?

Finally, the *Agree* group (17%) overwhelmingly either agreed or strongly agreed with all questions. This seems contradictory—for example, when a

respondent simultaneously strongly agrees that they don't have strong opinions about the collection and use of data about them and that they're concerned

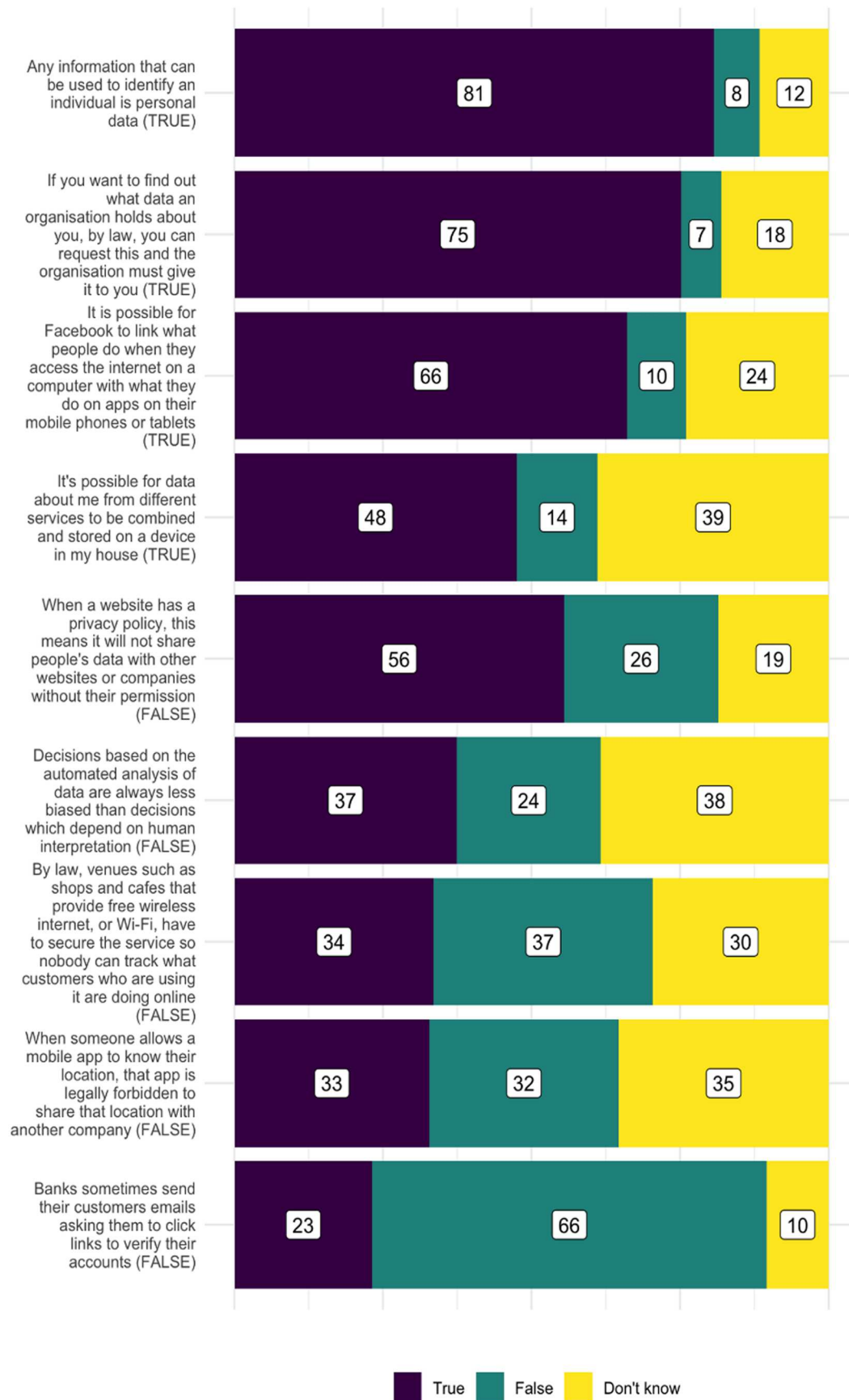


Figure 3. Are the following statements true or false?

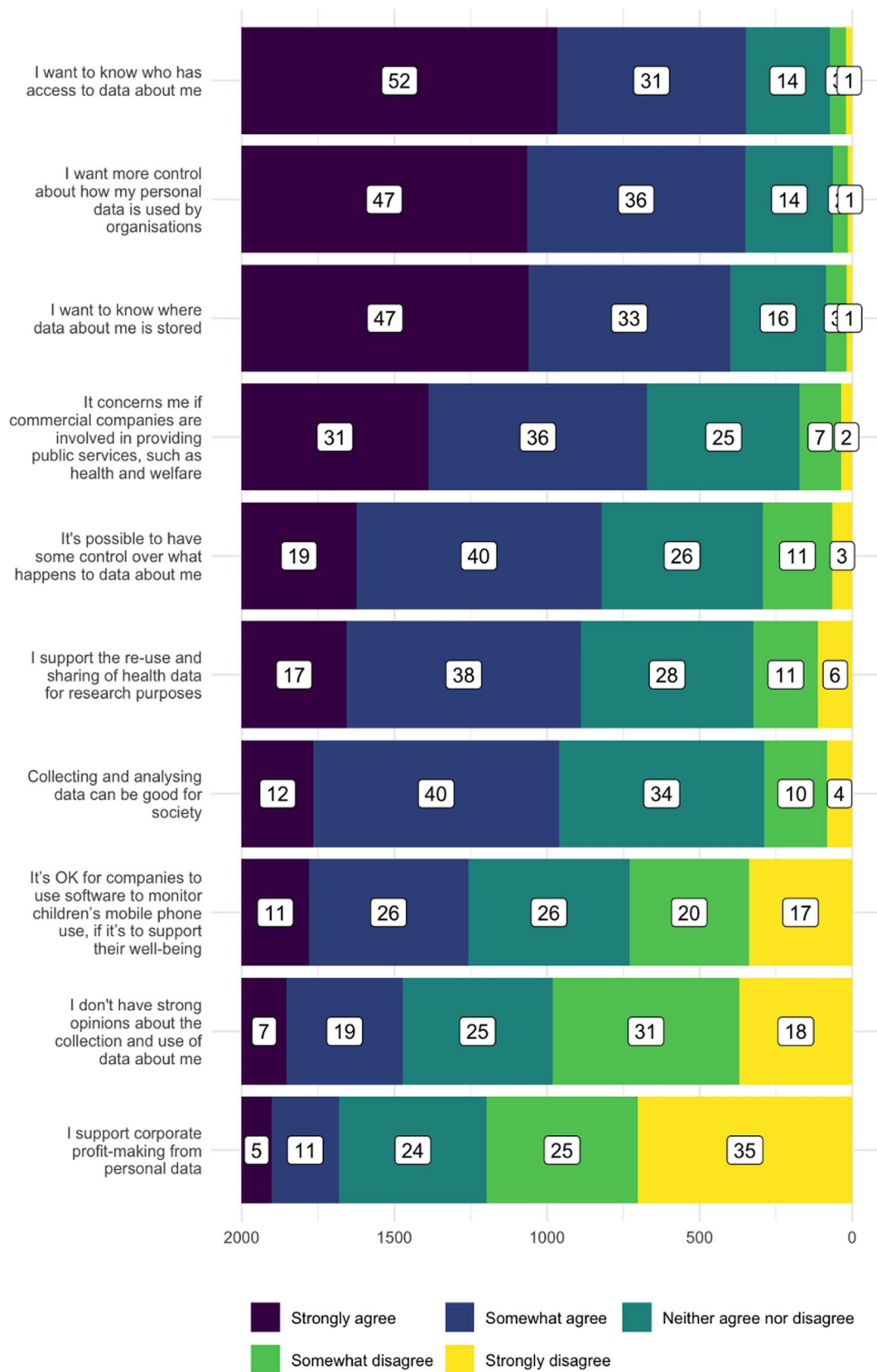


Figure 4. Please indicate how strongly you agree or disagree with the following statements.



**Table 2.** Understanding groups, attitudes groups, descriptions, and percentages of respondents in each group.

| Group name           | Group description                                                                                                                                    | % of respondents |
|----------------------|------------------------------------------------------------------------------------------------------------------------------------------------------|------------------|
| Understanding groups |                                                                                                                                                      |                  |
| Aware                | generally correctly identify true and false statements                                                                                               | 25               |
| Believers            | generally respond that the statements they are presented with are true                                                                               | 38               |
| Disbelievers         | generally respond that statements are false, even when they are in fact true                                                                         | 14               |
| Don't knows          | generally state that they don't know the answer to awareness questions                                                                               | 23               |
| Attitude groups      |                                                                                                                                                      |                  |
| Critical             | strongly disagree with some statements that are positive about data practices, and strongly agree with others that are negative about data practices | 36               |
| Cautious             | tend to agree or disagree in the same directions as the <i>Critical</i> group, but not strongly                                                      | 34               |
| Neutral              | overwhelmingly respond with "Neither agree nor disagree"                                                                                             | 13               |
| Agree                | overwhelmingly either agree or strongly agree with all statements                                                                                    | 17               |

about the role of commercial companies in public services. This suggests that no group fully accepts data uses, as even those who accept some data uses want information about and control over what happens to their data, as Figure 4 illustrates. This could be seen as “digital resignation” (Draper and Turow 2019) or reluctant acceptance of data uses (Peppin 2020). It is also possible that people within this group found the items difficult to interpret or that they used the “Strongly agree” category to communicate their general strength of feeling. Or they “satisfied” (Krosnick et al. 2001)—that is, they selected the box in the same place on each page to complete the survey quickly, something that is, of course, possible with all surveys.

Overall, what most distinguishes respondents in the attitudes latent classes is *strength* of feeling, regardless of domain, context, or other details about data uses. As latent class analysis involves bottom-up clustering of data, this finding is not an artifact of our methodological choices. We did not pre-define these clusters; rather, they are the clusters that best characterize patterns of responses. With that said, labeling involves choice. Our labels for latent classes are consistent with the patterns of responses that make them up, but other labels would be equally legitimate. We classify strong negative opinions about data uses as “critical,” where other terms, for example, “cynical,”

could also be used. It is also important to acknowledge the limitations of labels. For the awareness latent classes, we chose *Aware*, *Believers*, *Disbelievers*, and *Don't Knows*. But as people acquire their knowledge from a range of sources, including media reports which are not always factually accurate, some respondents in the *Disbelievers* latent classes may believe their answers to be accurate or correct, even if they are not, or they may have interpreted questions differently from those in the *Aware* class.

### **The relationship between awareness and attitudes**

Having established our latent classes, we then turned to consider the relationship between the two models, as illustrated in Figure 8. The bars connecting awareness latent classes to attitude latent classes show the relationship between them. The width of bars denotes the fraction of survey respondents who were members of both latent classes connected by the bars: the thicker the bar, the stronger the relationship.

The figure shows that 46% of *Aware* are in the *Critical* group, significantly more than the 36% of the overall sample who are *Critical*. They are also significantly less likely to be in the *Neutral* group, at 9% compared with 13% of the overall sample, and slightly less likely to be in the *Agree* group, at 14% compared with 17% of the overall sample.

*Believers* are more likely to *Agree* (22 vs. 17%) and less likely to be *Neutral* (9 vs. 13%) compared to the overall sample, but their proportions who are *Critical* and *Cautious* are very similar (37 vs. 36 and 33 vs. 34%, respectively). By contrast, the *Disbelievers* are the least likely to be *Critical* (20 vs. 36%), and the most likely to be either *Neutral* (21 vs. 13%) or *Agree* (25 vs. 17%). Finally, the *Don't know* group are moderately less likely to be *Critical* (34 vs. 36%) and more likely to be *Cautious* (41 vs. 34%), but also significantly more likely to be *Neutral* (19 vs. 13%) and less likely to be *Agree* (7 vs. 17%).

What is noteworthy here is that people who are most aware of data uses are also most critical or cautious about them. This finding challenges deficit model thinking which underlies data transparency and data literacy advocacy and initiatives and which has been criticized by PUS (Public Understanding of Science) scholars (Aitken, Cunningham-Burley, and Pagliari 2016; Irwin 2014). Such thinking and initiatives propose that the “data trust deficit” (RSS 2014) can be addressed with better information and better skills to understand information. Our findings suggest



**Figure 5.** Latent class model (awareness): first set of variables relating to awareness questions about data collection and use.

that it is not how much people know or are aware that matters, but rather the characteristics and degree of trustworthiness (O'Neill 2018) of the object of knowledge or awareness—in this case, data uses, which are perceived to be concerning. That the converse is also true reinforces this point—that is, the people who are the least aware of data uses (the *Disbelievers*) are also the least likely to be *Critical*, and most likely to be *Neutral*.

### **Differences in awareness and attitudes across demographic groups**

As well as establishing relationships between awareness and attitude latent classes, we explored whether there were differences in latent class membership by demographic group, given the role that structural inequalities can play in trust, acceptance, and other

attitudes. We focus on six demographic characteristics: gender, ethnicity, education, age group, disability, and sexuality. We do so because existing literature suggests that people in these demographic groups are more likely to be adversely affected by data uses than people in other groups and that this may affect awareness and attitudes (e.g. Benjamin 2014, 2016; Dobransky and Hargittai 2016; Kennedy, Steedman, and Jones 2020; Rendina and Mustanski 2018; Robinson et al. 2015; Woodruff et al. 2018).

Table 3 shows the differences in awareness of latent class membership by these demographic variables. Not all of these differences are statistically significant at the 95% level. For example, only around 5% of our sample are White Other, so a difference with another group of a few percentage points would not be statistically significant. In addition, while differences between individual characteristics may be statistically



**Figure 6.** Latent class model (awareness): second set of variables relating to true/false questions.

significant, they are not all jointly significant. That is, when two variables are controlled for, only one or neither of the differences is significant. For example, an observed difference in attitudes between age groups may be primarily explained by the fact that older people are less likely to have higher educational qualifications. Here, we highlight the differences that are statistically significant when modeled together using multinomial logistic regression.

As Table 3 shows, the largest differences relate to age. While all age groups are similarly likely to be *Aware*, people aged between 18 and 24 are particularly likely to be *Disbelievers*, and all other age groups are more likely to be *Believers*. People aged 55 and older are moderately more likely to be *Don't knows* than other age groups. There are also differences by ethnic group. Black and Asian people are significantly less likely to be in the *Aware* cluster and significantly

more likely to be either *Believers* or *Disbelievers*, although the difference in the probability of being *Disbelievers* is not statistically significant between White British and Black respondents.

In terms of gender differences, men are more likely to be in the *Knowledgeable* group, while women are more likely to be in the *Don't know* group. Men and women are similarly likely to be *Believers* and *Disbelievers*. Educational differences are also confined to these two latent classes. People with higher educational qualifications are more likely to be *Aware*, while people with fewer educational qualifications are more likely to be *Don't knows*. Differences between disabled and non-disabled people are small, whereas differences between LGBTQ+ and heterosexual cis-gender people are larger, as LGBTQ+ people are far more likely to be *Aware*, and far less likely to be *Believers*.

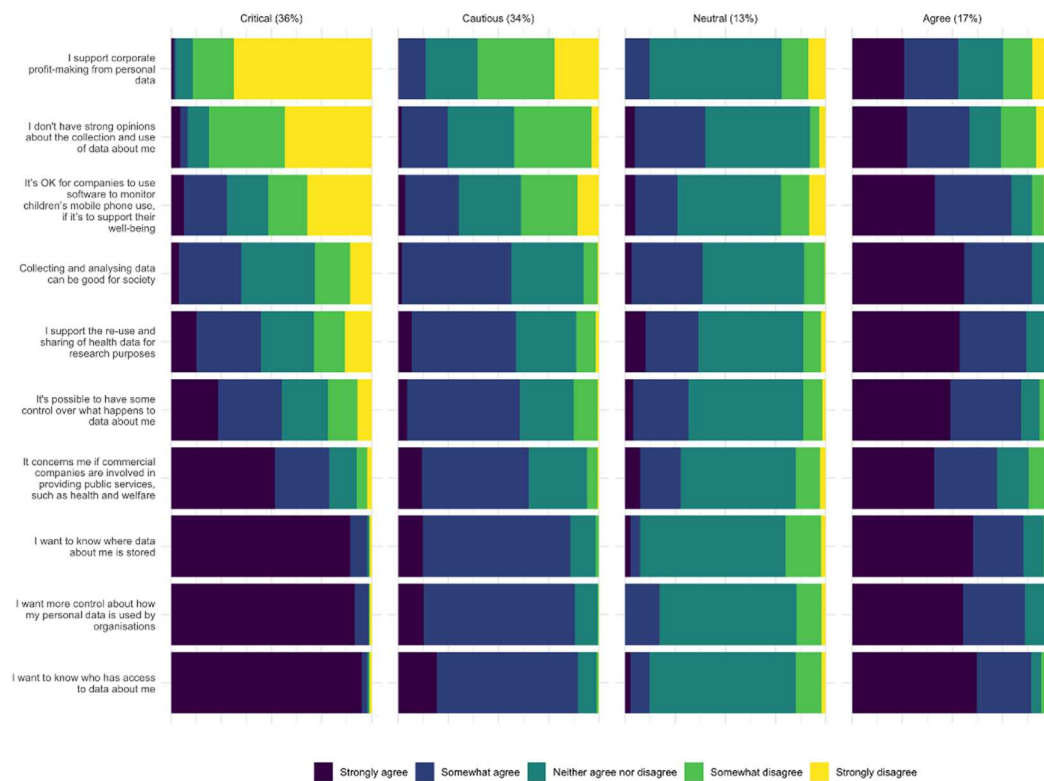
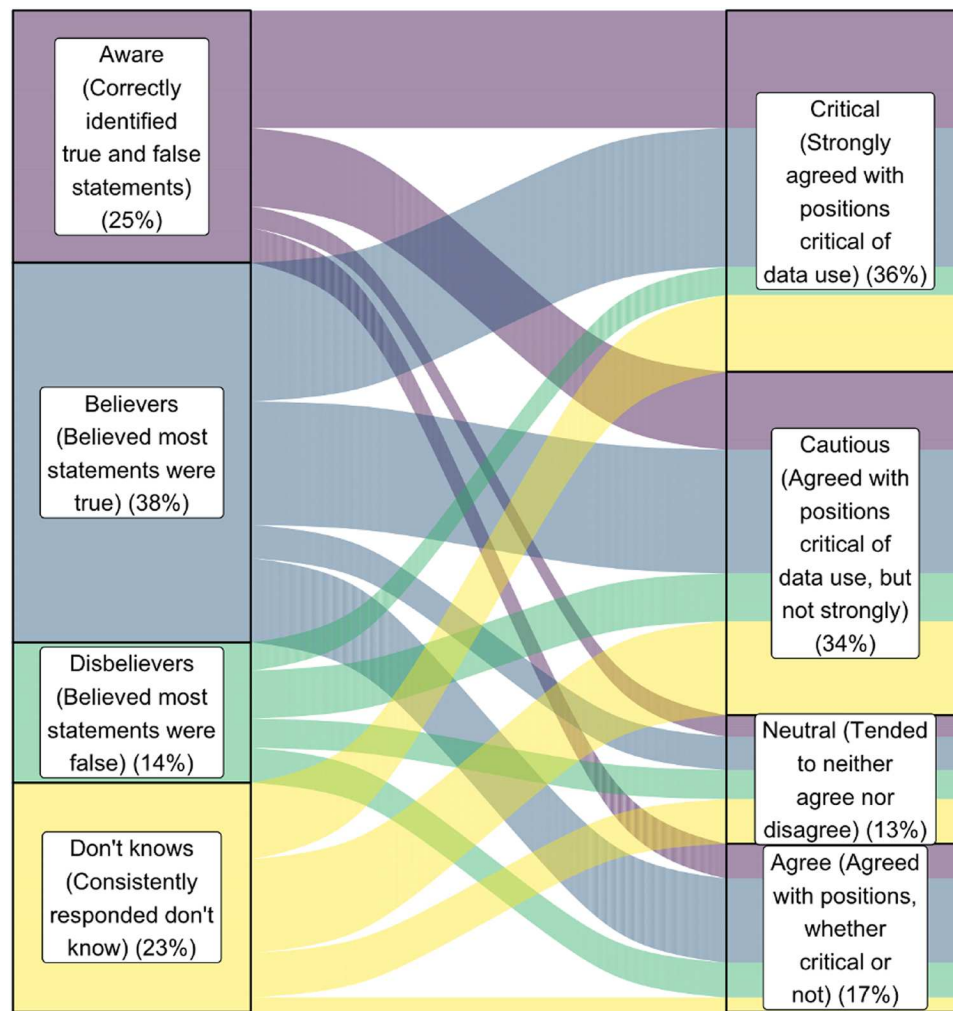


Figure 7. Latent class model (attitudes).

Table 4 shows the differences between these same groups in the attitude latent classes. Women are more likely to be in the *Critical* group than men, who are more likely to be *Neutral*. Differences in ethnic group are generally small, although Black people are less likely to be in the *Cautious* group and more likely to be in the *Neutral* group than are others. Differences by education are also generally small; while there are some differences (for example, people with qualifications taken around the age of 18, such as A-levels, are more likely to be in the *Cautious* group), it is likely that this is a difference that has emerged by chance because of the sampling process. Similarly, there are no significant differences between disabled people and non-disabled people, nor are there significant differences between LGBTQ+ people and heterosexual cisgender people. The largest differences are between different age groups. 27% of people in the youngest group are *Critical*, while 44% of people who are 55–64 and 42% of people aged 65+ are in this group. Sixteen to 24s are the most likely to be *Neutral* and older people are significantly less likely to be in the *Agree* latent class than others. Overall, other than between age groups, the differences between demographic groups are much smaller for the attitude latent classes than for the knowledge latent classes.

## Discussion and conclusions

Our survey confirms the finding of prior studies into people's awareness of data uses that people appear to be aware of certain data uses and less aware of others (e.g. Doteveryone 2018a). It also confirms the findings of research into attitudes, as it shows that public concern about certain data uses is high (ICO 2019; Pew 2019). Combining these two findings to explore the relationship between awareness and attitudes, we contribute original insights that advance understanding of public views on data uses. We found that people who are more aware of data uses tend to be more critical and cautious about them. This finding challenges the assumption that more and better information (as promoted by transparency advocates), combined with more and better skills to make sense of it (what data literacy initiatives seek to achieve) will result in greater acceptance of data uses. In our survey, the *Aware* latent class, or respondents who mostly answered the questions about data collection and use correctly and correctly identified the false statements in the true/false batch, were more likely to be in the *Critical* group, who strongly disagreed with negative statements and strongly agreed with positive statements, and moderately more likely to be in the *Cautious* group, who tended to agree or disagree with the same items.



**Figure 8.** The relationship between knowledge latent classes (on the left) and attitude latent classes (on the right).

We also contribute to debate about the role of structural inequalities in relation to perceptions of data uses. We found that respondents from different demographic groups are not equally likely to be in all latent classes. The main difference that we identified related to age. In the awareness classes, younger people were more likely to be *Believers*, while older people were more likely to be *Disbelievers*. With regard to attitudes to data uses, older people were more likely to be in the *Critical* latent class, while younger people were more likely to be in the *Neutral* class. There were other small differences across groups, but on the whole, we found different demographic groups have a lot in common in terms of attitudes to data uses, the unequal likelihood of being in latent classes notwithstanding. In our survey, it was not the case that people most likely to be negatively affected by data uses are the only ones who are concerned about them. Rather, people who do not belong to disadvantaged or minority groups are still worried

about how these groups might be negatively impacted by data uses. This finding that structural inequalities matter in perceptions of data uses, but not only to people whose lives are negatively affected by them, is an important contribution to data and inequalities literature.

One possible conclusion to draw from our finding—that people who are more aware of data uses are more critical and cautious about them—is that information about data uses should be made *less* available, to minimize caution, critique, and concern. Of course, that is not what we are suggesting. Our survey, like other surveys (e.g. Doteveryone 2018a, 2018b; Hartman et al. 2020), found that people want information about what happens to their personal data: they want to know who has access to data about them (83%) and where data about them is stored (80%). What is at stake here is not how much people know or are aware, but rather the characteristics of the object of knowledge or awareness.



**Table 3.** Awareness latent class membership by demographic variables.

| Variable     | Category                       | Aware (25%) | Believers (38%) | Disbelievers (14%) | Don't knows (23%) |
|--------------|--------------------------------|-------------|-----------------|--------------------|-------------------|
| Gender       | Male                           | 28%         | 38%             | 15%                | 20%               |
|              | Female                         | 23%         | 38%             | 13%                | 26%               |
| Ethnic group | White British                  | 26%         | 36%             | 12%                | 26%               |
|              | White other                    | 30%         | 41%             | 13%                | 16%               |
|              | Asian                          | 16%         | 49%             | 18%                | 16%               |
|              | Black                          | 19%         | 45%             | 21%                | 15%               |
|              | Mixed/other                    | 35%         | 29%             | 19%                | 17%               |
| Education    | No qualifications              | 17%         | 30%             | 18%                | 36%               |
|              | Vocational qualifications      | 18%         | 39%             | 18%                | 25%               |
|              | Qualifications taken around 16 | 20%         | 38%             | 12%                | 30%               |
|              | Qualifications taken around 18 | 26%         | 38%             | 13%                | 23%               |
|              | Some higher education          | 28%         | 37%             | 14%                | 20%               |
| Age group    | Degree                         | 29%         | 39%             | 14%                | 17%               |
|              | 18–24                          | 24%         | 27%             | 26%                | 22%               |
|              | 25–34                          | 27%         | 40%             | 17%                | 16%               |
|              | 35–44                          | 27%         | 38%             | 15%                | 19%               |
|              | 45–54                          | 24%         | 41%             | 10%                | 24%               |
|              | 55–64                          | 24%         | 39%             | 9%                 | 28%               |
| Disability   | 65 or older                    | 25%         | 38%             | 9%                 | 28%               |
|              | Disabled                       | 27%         | 33%             | 14%                | 26%               |
| Sexuality    | Not disabled                   | 25%         | 39%             | 14%                | 22%               |
|              | Heterosexual/straight          | 23%         | 39%             | 14%                | 23%               |
|              | LGBTQ+                         | 34%         | 31%             | 13%                | 22%               |

**Table 4.** Attitude latent class membership by demographic variables.

| Variable     | Category                       | Critical (36%) | Cautious (34%) | Neutral (13%) | Agree (17%) |
|--------------|--------------------------------|----------------|----------------|---------------|-------------|
| Gender       | Male                           | 33%            | 35%            | 15%           | 18%         |
|              | Female                         | 39%            | 34%            | 11%           | 16%         |
| Ethnic group | White British                  | 35%            | 36%            | 14%           | 15%         |
|              | White other                    | 41%            | 35%            | 9%            | 15%         |
|              | Asian                          | 35%            | 33%            | 12%           | 20%         |
|              | Black                          | 38%            | 20%            | 11%           | 31%         |
|              | Mixed/other                    | 40%            | 28%            | 11%           | 20%         |
| Education    | No qualifications              | 38%            | 28%            | 17%           | 18%         |
|              | Vocational qualifications      | 31%            | 30%            | 18%           | 20%         |
|              | Qualifications taken around 16 | 40%            | 33%            | 14%           | 13%         |
|              | Qualifications taken around 18 | 32%            | 39%            | 12%           | 16%         |
|              | Some higher education          | 37%            | 33%            | 13%           | 16%         |
| Age group    | Degree                         | 37%            | 35%            | 10%           | 18%         |
|              | 18–24                          | 27%            | 33%            | 18%           | 22%         |
|              | 25–34                          | 32%            | 31%            | 11%           | 25%         |
|              | 35–44                          | 35%            | 33%            | 11%           | 21%         |
|              | 45–54                          | 35%            | 37%            | 14%           | 14%         |
|              | 55–64                          | 44%            | 34%            | 15%           | 8%          |
| Disability   | 65 or older                    | 42%            | 38%            | 10%           | 11%         |
|              | Disabled                       | 37%            | 30%            | 13%           | 19%         |
| Sexuality    | Not disabled                   | 36%            | 36%            | 13%           | 16%         |
|              | Heterosexual/straight          | 36%            | 35%            | 13%           | 16%         |
|              | LGBTQ+                         | 36%            | 31%            | 13%           | 20%         |

As noted above, some researchers propose that attention should be focused on improving the quality and trustworthiness of technological or scientific systems, not public understanding of them. For example, Aitken, Cunningham-Burley, and Pagliari (2016, 713) argue that efforts to increase public understanding as a way of addressing attitudes like distrust are flawed, in part because the assumption that greater understanding of science results in more positive attitudes toward it “remains unproven.” Therefore they propose that the focus should not be on what the public understands, but rather on ensuring that science and

related systems are *trustworthy*. Similarly, writing about public attitudes to AI, Knowles and Richards (2021) argue that the public will trust AI when an effective regulatory ecosystem to oversee these technologies is put in place. They also suggest that where there is distrust or concern, it is the object of trust and its related regulatory framework that needs to change, for concern to diminish, not information about them or understanding of information (see also O'Neill 2018).

These writers are critical of the “deficit model” (Irwin 2014)—that is, the belief that the public has



gaps in its knowledge which, if filled, will lead to more positive attitudes—that underpin a misplaced focus on public trust rather than system trustworthiness. Likewise, Benjamin (2014, 2016) criticizes the assumption that “the transmission of information” about scientific systems will result in “the granting of permission” (2016, 967) or other forms of acceptance. She argues that such assumptions fail to acknowledge the ways in which structural inequalities unsettle such neat models. Our research shows that putting these ideas into dialogue with research into public views of data uses is productive. Despite longstanding criticisms, the deficit model continues to underpin policy, practitioner, and scholarly thinking, about public attitudes to data uses, as well as in other domains (Gorski 2011; Miles and Sullivan 2012). Better transparency and better literacy may improve understanding of data uses, and the former would certainly respond to our finding that people want more information about who has access to their data and where it is stored. However, given that people who are more aware of data uses are more critical and cautious about them, we argue that *better data uses* are essential if public attitudes to them are to improve—although what constitutes “better” data uses is, of course, open to debate.

As indicated by our use of the term *Aware* to describe one of our latent classes, we propose that what we have established is people’s awareness of data uses, with awareness understood as being conscious of something, and as generalized knowledge rather than full understanding of details. We found that respondents were more likely to state that false statements were true than vice versa, which could be seen as a reluctance to state categorically that data uses *do not* occur. Respondents stating that they didn’t know whether false statements were true or false could also be interpreted in the same way. Such reluctance may come from general awareness that data collection, analysis, and sharing take place, even if respondents are not certain about the particular data use mentioned in a statement. The forty-eight percent of respondents who believed organizations collect data by tracking people’s eye movements to track what they look at online—and members of our team who were also unsure about this—may also be reluctant to state categorically that it does not happen, because of awareness that it *may* happen.

If attitudes are understood as ways of perceiving the world around us that are generally consistent, and opinions as views or judgments about a particular phenomenon not necessarily based on fact or knowledge, we can characterize what we found as a mix of attitudes and opinions. Further, qualitative research is

needed to disentangle attitudes from opinions in our data. When respondents express concern about what happens to health data or companies monitoring children’s mobile phone use to support their well-being, it may be the data type, the people to whom the data refers, the purposes for which it is used, who is using it, the context, or other issues that concern them. To know whether these responses represent opinions—“views or judgments about a *particular* phenomenon”—or more generally consistent attitudes, it is necessary to ask *why* they have responded in the ways that they did, something that is most effectively done through qualitative research.

We have argued that it is important to reflect on the meanings of the terms we use in our claim-making and the methodological choices we make in our research design, including choices about question wording and the options we make available to respondents. The two paragraphs above represent our efforts to do so. By modeling a reflective lens, we advance scholarship on public views of data uses.

Reflecting on how to interpret expressions of uncertainty in survey research is also important. Our respondents were given don’t know options in response to awareness statements, and neither agree nor disagree options in response to attitude statements. What can we make of such responses? We found that there was a clear relationship between the two, between respondents who are *Don’t know*s in awareness and *Neutral* in attitude. A range of phenomena may motivate such answers. Some respondents will genuinely not know whether the statement they are being asked to assess is true or false, and have no view of the statements about which they are being invited to express opinions. Others may have some idea but not feel confident taking a position; others still may consider not taking a position to be the lowest-effort option available to them (Krosnick et al. 2001).

Our findings suggest some recommendations for data policy and practice. First, they indicate that if data uses remain unchanged, people are likely to continue to be concerned about them, despite transparent information or skills to interpret it. Certain data uses consistently concern people—as noted above, our research confirms the findings of previous studies in this regard. One example is sharing data originally intended for health or other prosocial purposes in ways that enable commercial companies to profit. Not engaging in such practices should therefore be considered.

Second, our findings indicated that while greater data transparency and data literacy initiatives are to

be encouraged, without changes to practice and process, people will continue to be concerned about data uses. We therefore argue that the focus of change should be on data uses, practices, and processes. Furthermore, it is important to consult the public continuously, particularly those negatively affected by data uses, about what changes they think are needed, because demographic groups differ in their attitudes to data uses. This will be resource-intensive, but engaging in an ongoing dialogue with the public about data uses on a case-by-case basis is necessary if data is to be used responsibly.

Finally, our findings indicate a need for data policymakers and practitioners to be wary of deficit model assumptions in their work. If "deficit ideology" (Gorski 2011) persists in approaches to public understanding of data uses, then the wrong things will be prioritized, as existing critiques of such ideology suggest. Unpacking assumptions about links between phenomena—in our case, attitudes to and awareness of data uses—is an important component of challenging deficit ideology.

Understanding people's awareness of and attitudes toward data uses is difficult for many reasons. Data uses are dynamic, they change constantly, and so does regulation, although legislation often struggles to keep pace. People's knowledge of and feelings about data uses are not static either—they can change in the process of thinking, reading, or talking about them. For these and other reasons, it is important not to oversimplify what it means to understand or be concerned about a data use. Likewise, it is important for public attitudes researchers to be explicit about how they define their chosen terms and reflexive about what their chosen methods do and do not enable them to make claims about. Our paper represents our attempt to do these things.

## Disclosure statement

No potential conflict of interest was reported by the author(s).

## Funding

This work was supported by the Nuffield Foundation under Grant [number OSP/43959].

## Data availability statement

Data are available through ORDA, the University of Sheffield's data repository, at doi: 10.15131/shef.data.c.6122043.v3.

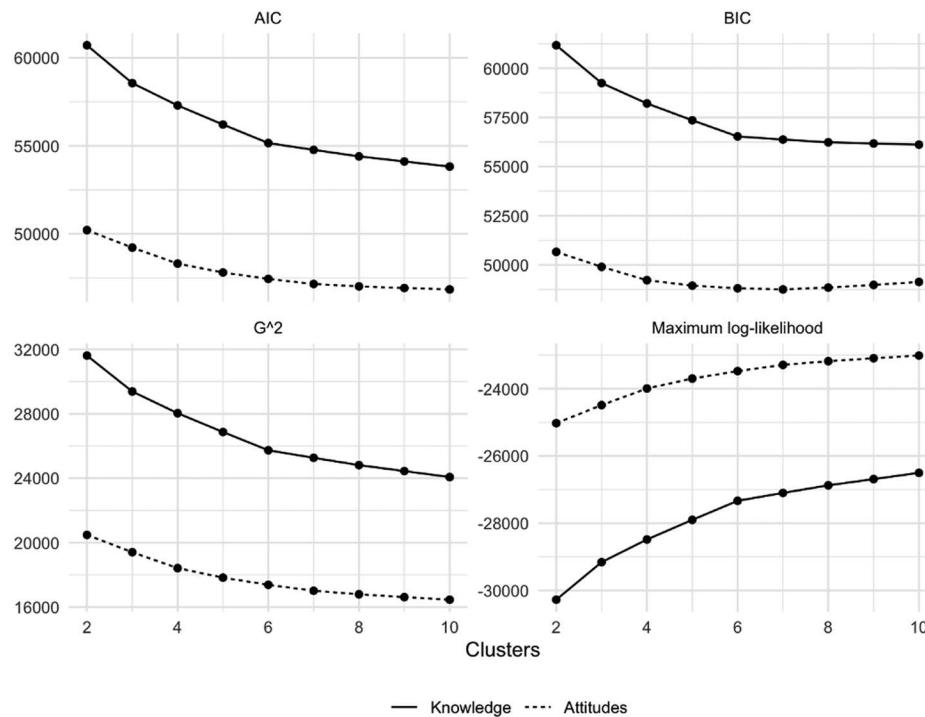
## References

- Ahmed, S. 2004. *The cultural politics of emotions*. Edinburgh: Edinburgh University Press.
- Aitken, M., S. Cunningham-Burley, and C. Pagliari. 2016. Moving from trust to trustworthiness: Experiences of public engagement in the Scottish Health Informatics Programme. *Science & Public Policy* 43 (5):713–23. doi: 10.1093/scipol/scv075.
- Ananny, M., and K. Crawford. 2018. Seeing without knowing: Limitations of the transparency ideal and its application to algorithmic accountability. *New Media & Society* 20 (3):973–89. doi: 10.1177/1461444816676645.
- Auxier, B., L. Rainie, M. Anderson, A. Perrin, M. Kumar, and E. Turner. 2019. Americans and privacy: Concerned, confused and feeling lack of control over their personal information. Accessed October 8, 2023. <https://www.pewresearch.org/internet/2019/11/15/americans-and-privacy-concerned-confused-and-feeling-lack-of-control-over-their-personal-information/>
- Bates, J., H. Kennedy, I. Medina Perea, S. Oman, and L. Pinney. 2023. Socially meaningful transparency in data-based systems: Reflections and proposals from practice. *Journal of Documentation*. doi: 10.1108/JD-01-2023-0006.
- Bem, D. J. 1970. *Beliefs, attitudes, and human affairs*. New York, NY: Brooks / Cole.
- Benjamin, R. 2014. Race for cures: Rethinking the racial logics of "trust" in biomedicine. *Sociology Compass* 8 (6):755–69. doi: 10.1111/soc4.12167.
- Benjamin, R. 2016. Informed refusal: Towards a justice-based bioethics. *Science, Technology and Human Values* 41 (6):967–90. doi: 10.1177/0162243916656059.
- Bergstrom, A. 2015. Online privacy concerns: A broad approach to understanding the concerns of different groups for different uses. *Computers in Human Behavior* 53:419–26. doi: 10.1016/j.chb.2015.07.025.
- Bucher, T. 2017. The algorithmic imaginary: Exploring the ordinary affects of Facebook algorithms. *Information, Communication & Society* 20 (1):30–44. doi: 10.1080/1369118X.2016.1154086.
- CDEI (Centre for Data Ethics and Innovation). 2021. Complete transparency, complete simplicity: How can the public sector be meaningfully transparent about algorithmic decision making? Accessed October 8, 2023. <https://www.gov.uk/government/publications/cdei-published-commissioned-research-on-algorithmic-transparency-in-the-public-sector>
- D'Ignazio, C. 2017. Creative data literacy: Bridging the gap between the data-haves and data-have nots. *Information Design Journal* 23 (1):6–18.
- Department for Digital, Culture, Media and Sport. 2020. National data strategy. Accessed November 3, 2023. <https://www.gov.uk/government/publications/uk-national-data-strategy/national-data-strategy>
- Ditchfield, H., H. Kennedy, S. Oman, L. Pinney, J. Bates, M. Fratzak, I. Medina-Perea, and M. Taylor. 2002. Research on *Living With Data* interviews & focus groups. Accessed November 7, 2023. <https://livingwithdata.org/project/wp-content/uploads/2022/07/living-with-data-qualitative-research-report-july-2002.pdf>

- Dobransky, K., and E. Hargittai. 2016. Unrealized potential: Exploring the digital disability divide. *Poetics* 58:18–28. doi: 10.1016/j.poetic.2016.08.003.
- Doteveryone. 2018a. People, power, and technology: The 2018 digital understanding report. Accessed October 22, 2023. <https://understanding.doteveryone.org.uk>
- Doteveryone. 2018b. People, power and technology: The 2018 digital attitudes report. Accessed October 22, 2023. <https://understanding.doteveryone.org.uk>
- Doyle McCarthy, E. 1996. *Knowledge as culture: The new sociology of knowledge*. Basingstoke, UK: Taylor and Francis.
- Draper, N. A., and J. Trow. 2019. The corporate cultivation of digital resignation. *New Media & Society* 21 (8):1824–39. doi: 10.1177/1461444819833331.
- EC (European Commission). 2019. Special Eurobarometer 487a. Summary – The General Data Protection Regulation. Accessed October 22, 2023. <https://cnpd.public.lu/content/dam/cnpd/fr/actualites/international/2019/ebs487a-GDPR-sum-en.pdf>
- Edwards, R., V. Gillies, and S. Gorin. 2021. Is breakdown in trust, transparency and social cohesion a price worth paying for more extensive data linkage? Accessed October 22, 2023. <https://blogs.lse.ac.uk/impactofsocialsciences/2021/05/06/is-a-breakdown-in-trust-transparency-and-social-cohesion-a-price-worth-paying-for-more-extensive-data-linkage/>
- Eslami, M., A. Rickman, K. Vaccaro, A. Aleyasen, A. Vuong, K. Karahalios, K. Hamilton, and C. Sandvig. 2015. “I always assumed that I wasn’t really that close to [her]”: Reasoning about invisible algorithms in news feeds. In *CHI ’15: Proceedings of the 2015 CHI Conference on Human Factors in Computing Systems*, 153–62. New York, NY: ACM.
- European Union. 2016. General data protection regulation. *Official Journal of the European Union* 49:L119.
- Factconference.org. 2018. FAT\*: 2018 Call for papers. Accessed June 5, 2022. <https://factconference.org/2018/cfp.html>
- Freire, P. 1996. *Pedagogy of the oppressed*. London: Penguin.
- Gelman, A., S. Goel, D. Rothschild, and W. Wang. 2016. High-frequency polling with non-representative data. In *Political communication in real time: Theoretical and applied research approaches*, eds. D. Schill, R. Kirk, and A. E. Jasperson, 89–105. London: Taylor and Francis.
- Gorski, P. 2011. Unlearning deficit ideology and the scornful gaze: Thoughts on authenticating the class discourse in education. *Counterpoints* 402:152–73.
- Gray, J., C. Gerlitz, and L. Bounegru. 2018. Data infrastructure literacy. *Big Data & Society* 5 (2):205395171878631. doi: 10.1177/2053951718786316.
- Hartman, T., H. Kennedy, R. Steedman, and R. Jones. 2020. Public perceptions of good data management: Findings from a UK-based survey. *Big Data & Society* 7 (1):205395172093561. doi: 10.1177/2053951720935616.
- ICO (Information Commissioner’s Office). 2019. Information commissioner’s annual report and financial statements 2018–19. Accessed October 22, 2023. <https://ico.org.uk/media/about-the-ico/documents/2615262/annual-report-201819.pdf>
- IOMC. n.d. Perception. Accessed October 22, 2023. <https://www.iomcworld.org/medical-journals/perception-39571.html>
- Irwin, A. 2014. From deficit to democracy (re-visited). *Public Understanding of Science* 23 (1):71–6. doi: 10.1177/0963662513510646.
- Kennedy, H., R. Steedman, and R. Jones. 2020. Approaching public perceptions of datafication through the lens of inequality: A case study in public service media. *Information, Communication & Society* 24 (12):1745–61. doi: 10.1080/1369118X.2020.1736122.
- Kennedy, H., S. Oman, M. Taylor, J. Bates, and R. Steedman. 2020. Public understanding and perceptions of data practices: A review of existing research. Accessed October 22, 2023. <https://livingwithdata.org/project/wp-content/uploads/2020/05/living-with-data-2020-review-of-existing-research.pdf>
- Knowles, B., and J. T. Richards. 2021. The sanction of authority: Promoting public trust in AI. In *FACCT ’21: Proceedings of the 2021 ACM Conference on Fairness, Accountability, and Transparency*, 262–71. New York, NY: ACM.
- Krosnick, J. A., A. L. Holbrook, M. K. Berent, R. T. Carson, W. Michael Hanemann, R. J. Kopp, R. Cameron Mitchell, S. Presser, P. A. Ruud, V. Kerry Smith, W.R. Moody, M. C. Green, and M. Conaway. 2001. The impact of “no opinion” response options on data quality: Non-attitude reduction or an invitation to satisfice? *Public Opinion Quarterly* 66 (3):371–403. doi: 10.1086/341394.
- Lupton, D., and M. Michael. 2017. “Depends on who’s got the data”: Public understandings of personal digital dataveillance. *Surveillance & Society* 15 (2):254–68. doi: 10.24908/ss.v15i2.6332.
- Miles, A., and A. Sullivan. 2012. Understanding participation in culture and sport: Mixing methods, reordering knowledges. *Cultural Trends* 21 (4):311–24. doi: 10.1080/09548963.2012.726795.
- Myers, D. 2012. *Social psychology*. New York, NY: McGraw-Hill.
- O’Neill, O. 2018. Linking trust to trustworthiness. *International Journal of Philosophical Studies* 26 (2):293–300. doi: 10.1080/09672559.2018.1454637.
- Oman, S. 2021. *Understanding well-being data: Improving social and cultural policy, practice and research*. Basingstoke, UK: Palgrave Macmillan.
- Oman, S. 2022. Doing understanding differently: Rethinking our relationship with metrics – And the people behind the data. *The Sociological Review Magazine*, July. Accessed October 22, 2023. <https://thesociologicalreview.org/magazine/july-2022/metrics/doing-understanding-differently/>
- Oskamp, S., and P. W. Schultz. 2014. *Attitudes and opinions*. Basingstoke, UK: Taylor and Francis.
- Pasquale, F. 2015. *The black box society: The secret algorithms that control money and information*. Cambridge, MA: Harvard University Press.
- Pew Research Center - Brooke Auxier, Lee Rainie, Monica Anderson, Andrew Perrin, Madhu Kumar & Erica Turner. 2019. Americans and Privacy: Concerned, Confused and Feeling Lack of Control Over Their Personal Information. <https://www.pewresearch.org/internet/2019/11/15/americans-and-privacy-concerned-confused-and-feeling-lack-of-control-over-their-personal-information/>. Accessed 20 November 2023.
- Peppin, A. 2020. Do we reluctantly accept digital tech? Accessed October 22, 2023. <https://medium.com/swlh/do-we-reluctantly-accept-digital-tech-9c91cff73657>
- Rendina, H. J., and B. Mustanski. 2018. Privacy, trust, and data sharing in web-based and mobile research: Participant perspectives in a large nationwide sample of men who have sex with men in the United States. *Journal of*

- Medical Internet Research* 20 (7):e233. doi: [10.2196/jmir.9019](https://doi.org/10.2196/jmir.9019).
- Robinson, G., and H. Dolk. 2015. Research update: Public attitudes to data sharing in Northern Ireland. Accessed October 22, 2023. [https://ark.ac.uk/ARK/sites/default/files/2018-07/update108\\_0.pdf](https://ark.ac.uk/ARK/sites/default/files/2018-07/update108_0.pdf)
- Robinson, L., S. R. Cotten, H. Ono, A. Quan-Haase, G. Mesch, W. Chen, J. Schulz, T. M. Hale, and M. J. Stern. 2015. Digital inequalities and why they matter. *Information, Communication & Society* 18 (5):569–82. doi: [10.1080/1369118X.2015.1012532](https://doi.org/10.1080/1369118X.2015.1012532).
- RSS (Royal Statistical Society). 2014. Royal Statistical Society research on trust in data and attitudes toward data use/data sharing. Accessed November 7, 2023. <https://www.ipsos.com/en-uk/new-research-finds-data-trust-deficit-lessons-policy-makers>
- Sander, I. 2021. Critical (big) data literacy. Accessed October 22, 2023. [https://www.bigdataliteracy.net/glossar/?rdp\\_we\\_resource=https%3A//www.ds-glossar.de/wiki/Critical\\_Big\\_Data\\_Literacy\\_\(Medienbildung\)](https://www.bigdataliteracy.net/glossar/?rdp_we_resource=https%3A//www.ds-glossar.de/wiki/Critical_Big_Data_Literacy_(Medienbildung))
- Woodruff, A., S. E. Fox, S. Rousso-Schindler, and J. Warshaw. 2018. A qualitative exploration of perceptions of algorithmic fairness. In *CHI '18: Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems*, 1–14. New York, NY: ACM. doi: [10.1145/3173574.3174230](https://doi.org/10.1145/3173574.3174230).
- Yates, S., E. Carmi, E. Lockley, B. Wessels, and A. Pawluczuk. n.d. Me and my big data: Understanding citizens data literacies research report. Accessed October 22, 2023. [https://www.liverpool.ac.uk/media/livacuk/humanitiesampsocialsciences/meandmybigdata/Understanding\\_Citizens\\_Data\\_Literacies\\_Research\\_Report\\_Final.pdf](https://www.liverpool.ac.uk/media/livacuk/humanitiesampsocialsciences/meandmybigdata/Understanding_Citizens_Data_Literacies_Research_Report_Final.pdf)

## Appendix A



**Figure A1.** Goodness-of-fit statistics for latent class analysis for questions on both knowledge and attitudes with numbers of clusters varying from 2 to 10.