Transportation Research Interdisciplinary Perspectives 24 (2024) 101022

Contents lists available at ScienceDirect

Transportation Research Interdisciplinary Perspectives

journal homepage: www.sciencedirect.com/journal/transportation-research-interdisciplinary-perspectives

Privacy preferences in automotive data collection

Anna Dowthwaite[a,*], Dave Cook[b], Anna L. Cox[a]

[a] UCL Interaction Centre, United Kingdom
[b] UCL Anthropology, United Kingdom

ARTICLE INFO

Keywords:
Connected cars
Privacy
Automotive data
Human-data interaction
Data protection
Data disclosure behaviour

ABSTRACT

Connected cars are becoming commonplace, creating vast volumes of data that may contain or reveal information about drivers. It is imperative to understand drivers' perspectives on such data being collected and used by car manufacturers. Applying the Human-Data Interaction (HDI) framework - which centres the user and their experience - to this context, we conducted semi-structured interviews with 15 drivers. Interview transcripts revealed issues with understanding of car data (Legibility) and drivers' sense of control over automotive data (Agency), across different circumstances (Negotiability). Our findings suggest that car manufacturers should enable learning, access, and control over car data via the mobile app in a coordinated fashion, as the privacy preferences of drivers are often based on perceived benefit or threat resulting from data collection. The ability to set data-sharing preferences in a time- and location- sensitive manner can help drivers navigate data sharing consent based on circumstances. These findings have implications for the consent procedures in modern cars as well as for the development of data-sharing programmes aimed at creation of climate-smart cities.

1. Introduction

Connected cars are becoming commonplace in the automotive sector and are creating vast volumes of data. The car manufacturers' ability to collect data will only increase as the cars they produce become more connected. The consultancy McKinsey predicts that, by 2030, 95 % of new vehicles will be connected, with nearly half of those having intermediate or advanced connectivity (Bertoncello et al., 2021). Car manufacturers and their potential partners could monetise these, often sensitive, data in ways which users may feel are inappropriate and which could potentially infringe on their privacy preferences. Connected cars that increasingly gather large volumes of data present a particular challenge: while a typical person can stop using social media or switch to a different search engine, many cannot avoid using their cars. This is especially the case for people with disabilities and those with dependants or without access to reliable public transport. It is therefore imperative to understand drivers' perspectives on such data being collected and used by car manufacturers.

Existing research in this area has focused on two, commercial-centric, strands of investigation. The first strand pertains to technology-driven solutions supporting data privacy and security for connected car networks (see Ram, Markkula, Friman, & Raz, 2018), for example so that cars cannot be hacked. The second strand centres on overcoming adoption barriers, be it by examining users' attitudes and preferences (Gurumurthy & Kockelman, 2020) or highlighting privacy issues that could jeopardise adoption of autonomous vehicles (Kaur & Rampersad, 2018). Some technical solutions have been proposed, for example, interfaces (Walter, Abendroth, & Agarwal, 2017) that, if employed by car companies, could help users control their own data. However, such research is rare and not enough attention is given to how users feel about the increasing datafication of the car sector and the ways that this could affect them as drivers and car owners.

In this paper we report the application of the Human-Data Interaction (HDI) framework - which centres the user and their experience – to this context. We conducted semi-structured interviews with 15 drivers in order to investigate how collection of data from connected cars, and the ways such data are or could be utilised by car manufacturers, are experienced by drivers. The key contributions of this work are:

(1) The first application of the Human-Data Interaction framework to connected car data.
(2) Demonstration of the usefulness of the Human-Data Interaction framework for the automotive sector.
(3) Design recommendations for designing data consent procedures in connected cars.

* Corresponding author at: UCL Interaction Centre, 66-72 Gower St, London, WC1E 6EA, United Kingdom
E-mail address: anna.rudnicka.15@ucl.ac.uk (A. Dowthwaite).

https://doi.org/10.1016/j.trip.2024.101022
Received 13 September 2023; Received in revised form 9 January 2024; Accepted 26 January 2024
Available online 13 February 2024

This work is of high importance for the emerging relationship between drivers as data subjects and car companies as data collectors and controllers. Our work provides a foundation for future investigations, by demonstrating how drivers make sense of data, what their priorities are and what they expect from car manufacturers in relation to data protection and privacy.

2. Related work

2.1. A moving panopticon

Data are crucial for modern cars. For connected cars, collection of data enables a range of connectivity features, from GPS navigation to 'automated driving' features in the case of newer car models. As highlighted by Jakobi et al., most people cannot avoid using cars (Jakobi, Alizadeh, Marburger, & Stevens, 2021). The increasing collection of data through cars is therefore a privacy issue that affects a large swathe of the society. While, in a digitised world, data protection and privacy present a challenge for most areas of life, the enclosed nature of a car's cabin is reminiscent of Foucault's panopticon, a space where one is unable to escape surveillance (Eski & Schuilenburg, 2022).

There are trends in the automotive industry that signal a potential future of business models reliant, at least partially, on surveillance, with data collected by cars becoming another source of income for car manufacturers. With companies like the Swedish telecommunications giant Ericsson or the Israeli start-up Otonomo helping manufacturers leverage car data, the focus is increasingly on ensuring that cars can generate profits throughout their lifecycle. While some profit can be gained from the sale of extended connectivity features (e.g., automated driving), data could also be re-sold to stakeholders like insurers or advertisers.

This is in line with Shoshana Zuboff's warning about our society's over-reliance on surveillance capitalism, an economic system underpinned by tracking behaviour. The beginnings of surveillance capitalism took place without much oversight from the consumer or the regulator (Zuboff, 2019). It is noteworthy that the changes in how cars are designed and monetised are developing with more awareness on the part of the public. Cars currently on the roads are collecting vast amounts of data, including deeply personal information about users' whereabouts, travel companions as well as data on other road users (Hajlaoui, Moulahi, & Guyennet, 2019). This is increasingly being highlighted in popular press, as drivers are warned about the scope of the information that their cars collect about them (Fowler, 2019). The next decade is likely to determine whether cars will remain a sphere of personal privacy (Walter, 2018) or become a mobile electronic Panopticon (Lyon, 1993).

Nevertheless, privacy research around connected and automated cars rarely explores the data protection responsibilities of manufacturers. Instead, the focus of manuscripts concerned with connected cars is predominantly commercial-centric. First, there is a wealth of research exploring connected car privacy from the standpoint of security and protection of users from bad actors. Having conducted a systematic mapping study that examined studies exploring either security or privacy in connected cars, Ram et al. (2018) posited that privacy-specific concerns may result from insecure infrastructure or an insecure system design. Indeed, many other researchers focus on 'solving' privacy issues in the area of modern cars by developing technical solutions. Recent examples include a framework for detection of vehicular bots (Rahal et al., 2022), a blockchain-based authentication mechanism (Zhang & Wu, 2021), and a method that uses dedicated short-range communication to enable communication between vehicles but obscure it from the server (Lim, Kim, Yu, & Lee, 2020).

Furthermore, a popular researchers' focus has been the examination of connected car privacy as a barrier to product adoption, for example through a privacy calculus perspective (e.g. Buck & Reith, 2020), with Derikx et al. (Derikx, De Reuver, & Kroesen, 2016) suggesting that privacy concerns in this context can be alleviated by offering consumers an insurance discount. Nevertheless, research suggesting that consumers may not have a full view of what data are collected via their cars (Frassinelli, Park, & Nürnberger, 2020) necessitates scepticism when assuming acceptance of car data collection on the part of users and situating it within a particular (monetary) value structure.

There is a need to explore users' experiences and attitudes to data privacy in increasingly connected modern cars, employing more user-centric methodology. This is important in order to create sustainable guidelines for designers that lead to technology solutions that really address the users' needs and to inform public policy on data protection for the mobility sector.

2.2. De-identification of car data

The Ontario Privacy Commissioner defined de-identification as "the general term for the process of removing personal information from a data set". De-identification of datasets is seen to protect the privacy of individuals (Information and Privacy Commissioner of Ontario, 2016). De-identification is a complex concept and process, however, and there is no absolute standard against which a dataset can be considered to be de-identified: there is a degree of subjectivity involved in considering whether a dataset has been adequately de-identified. The Ontario Privacy commissioner, for example, states that data need to be removed where there is a 'reasonable expectation' that the information could be used to identify the individual. De-identified data could be subject to re-identification 'attacks' where an attempt is made, potentially using other information, to identify an individual.

There are various approaches that can be taken to de-identification and Löbner et al. (2021) list eight major classes of de-identification techniques: (1) statistical tools (e.g., sampling and aggregation), (2) cryptographic tools (e.g., deterministic, order-preserving, Homomorphic Encryption, secret sharing), (3) suppression (e.g., masking, local suppression, record suppression, sampling), (4) pseudonymization, (5) granularization (reducing the granularity of information for example by rounding or top/bottom coding), (6) randomization through random modification of attributes for example by noise addition, permutation, or micro aggregation, (7) differential privacy, which involves 'sharing information in a dataset while withholding information about a single information in that dataset', and (8) k-anonymity 'defining a state where a person cannot be distinguished from $k-1$ other persons in a dataset' (Löbner et al., 2021; Samarati & Sweeney, 1998).

We can give an example that makes clear how de-identification of travel data in the context of cars might work in practice. The location of the vehicle over time is a variable that can be used to re-identify users, even if data that directly identify the users have been removed. Re-identification might take place, for example, by using these data to find the user's home address and hence their name.

One approach would be to mask location data, either by its complete removal or by replacing it with an encrypted version. If this is not possible then it might be generalised by making the location data less precise so that using the data to tie the user to a precise home address is not possible. This example shows how de-identification is not an absolute step. Here, generalised data, combined with other data about the user, such as age or occupation could be used to identify the user from a small group of individuals in a given neighbourhood.

Furthermore, it is crucial to consider privacy of drivers and passengers in a context of data use that is broader than the current use of car data or the current capabilities of de-identification.

With the popularisation of AI tools and increasing access to these, we must consider the potential future implications of the depth of surveillance enabled through connected car technology. Even if the extensive data gathered on drivers and passengers can be kept private and only used for legitimate purposes at this time, these datasets could be used for nefarious purposes in the future as de-identification methods evolve. In the following section we discuss the potential avenues for the misuse of car data.

2.3. Misuse of car data

One important context for users considering whether to share their car data (e.g., by navigating privacy settings in their car or car-associated app) is that users may have concerns about potential for unethical, exploitative or even criminal use of that data. Misuse of car data could happen at different stages of data handling - one needs to consider not only the cybersecurity and ethics of data use within car companies but also the way data are used and handled. Some use cases might cause drivers some concerns, and we discuss these below.

Car companies have a record of questionable behaviour. The Volkswagen emissions scandal saw the company programming emissions control systems on 11 million cars so as to generate misleading data during testing (Jung & Sharon, 2019). As a result, the company has been subject to extensive regulatory and legal action resulting in billions of dollars in damages. This case provides an example of how the data collected by cars, and controlled by car manufacturers, might be of great interest to both drivers and the society at large.

Collection and misuse of car data has not, so far, led to a scandal of comparable scale, but customers are unlikely to be reassured either by car companies' ethical track records or by the care they take with customer data. One piece of evidence for the lack of care with customer data shown by car makers is their extensive sharing of car data with third party brokers. These brokers (known as vehicle data hubs) will combine car data with data from other sources. This may be used to de-anonymise aggregated data supplied by the car company. This, by now, valuable, highly sensitive and personal data can then be sold to third parties (Caltrider, Rykov, & MacDonald, 2023). This may lead to marketing of products or services that the user may regard as intrusive given it shows that the advertiser has deep and extensive knowledge of their lifestyle.

A more direct example of why it is important for drivers to understand what data are collected by their car and how to keep these data secure is domestic abuse, with reports of abusers using proprietary applications to track survivors' location in the car in real-time (Stephenson et al., 2023).

Finally, if car data fall into the hands of a criminal third party through a data breach, then an even wider range of criminal and unethical uses of the data becomes possible. Users are unlikely to take comfort from the quality of security measures taken by leading car companies. As reported by the Mozilla Foundation who reviewed privacy and security of 25 leading car brands, 17 brands had a bad track record, with multiple leaks and security breaches (Caltrider et al., 2023).

2.4. From data privacy to Human-Data Interaction

The first academic definition of privacy refers to the right to be left alone (Warren & Brandeis, 1989) which was later seen as giving grounds for protection against intrusion, appropriation, defamation and (most relevant to this paper) unwanted disclosure (Richards & Solove, 2010).

Current research perspectives on privacy are varied and can be conflicting (Iachello & Hong, 2007). Much of the research focused on data privacy looks at the so-called privacy paradox, whereby users hold privacy concerns but do not act on them or act in contradiction to them – a phenomenon often explained by the existence of a privacy calculus whereby consumers disclose information because they perceive the benefits of disclosure to outweigh the risks (Kokolakis, 2017). However, recently scholars have argued that we need to go beyond the privacy paradox and be more sensitive to user needs and expectations in a particular context (Nissenbaum, 2004). Zuboff has written extensively about 'surveillance capitalism', an economic reality that makes it near impossible for consumers not to disclose their data (Zuboff, 2019). Indeed, Martin (Martin, 2020) has warned that viewing privacy from the standpoint of the privacy paradox can serve to legitimise a status quo in which organisations do not seek to understand and/or respects users' reasonable privacy expectations and preferences.

What makes privacy in cars a particularly complex issue is that the extent of collection of car data (here referred to broadly as all data collected via the car, for example through internal and external car sensors) is often unclear, with reverse-engineering research demonstrating that even older models of cars collect a surprising range of data (Jakobi et al., 2021). It therefore follows that merely assessing the privacy concerns that users do or may have in relation to increasingly connected cars, without discussing their understanding of what data the car collects, may not present us with the full picture of how users experience and what they expect regarding data privacy in cars.

Human-Data Interaction [HDI], a framework parallel to the field of Human-Computer Interaction, suggests that it is important to study data privacy topics in a more user-centric way, by studying the relationship between users and data. Mortier et al. (Mortier, Haddadi, Henderson et al., 2014) explain the framework in relation to three principles one ought to focus on when exploring the impact of new technologies which involve interaction of users with data: *Legibility*, *Agency* and *Negotiability*. These correspond with, respectively, the users' ability to understand, control and negotiate across contexts what happens to their data. Although the HDI lens has been extensively used in health informatics and has also helped researchers explore data ownership in smart cities (Victorelli, Dos Reis, Hornung, & Prado, 2020), it has not, to the best of our knowledge, been applied to the area of connected cars.

The Human-Data Interaction framework is particularly appropriate for examination of privacy issues in connected cars because it brings into sharp focus the relationship between data and the self. The complexity of associated technologies and the fast pace of change in the industry could result in the exclusion of the user from decision-making about what data are collected via their cars and how they can be used. This suggests a need for privacy research in the connected car context that centres the user and their experiences and preferences.

3. Current research

In this study, we explored how drivers experience and think of the data their cars collect, by employing the lens of the Human-Data Interaction framework (Mortier et al., 2014). This framework focuses on three principles, Legibility, Agency, and Negotiability.

In this context of connected cars, we saw *Legibility* as being able to understand what data are collected by a car and what happens to them later. *Agency* was seen as ability to consent to the collection and use of such data by car companies and ability to access and control data after they have been collected. *Negotiability* was seen to refer to how drivers' opinions may shift over time and in different contexts.

Although during interviews we explored drivers' views on both the commercial and public sector's use of car data, this study explored, specifically, how drivers experience and think of data collection in the context of their relationships with their car manufacturers. Since one of the most popular manufacturers of highly connected vehicles, Tesla, offers car insurance, the use of data for insurance purposes was also considered to be within the scope of our investigation. We set out to answer an overarching research question:

RQ: What are connected car drivers' experiences of, and opinions about, the collection and use of car data by car companies?

We aimed to identify the key areas of importance and concern for drivers in the context of car connectivity, enabling us to provide design recommendations for supporting Legibility and Negotiability of car data and the Agency of drivers.

4. Method

4.1. Participants

Participants were recruited via social media, word of mouth and through snowballing. All participants had to meet the following criteria: (1) 18 years of age, (2) resident in the United Kingdom, (3) owner of a car with connectivity features. Recruitment messages on Twitter were amplified through paid advertisements. To streamline recruitment, we identified the UK brands across which all car makes were connected at the time of recruitment: Tesla, Mini, Jaguar Land Rover, Audi, Volvo, Mercedes, and BMW. These brands were then highlighted in recruitment messages. We also recruited an additional participant, a Kia driver, having confirmed that their car model had connectivity features. We aimed to conduct a total of 15 interviews and slots were assigned on a 'first come first serve' basis to eligible individuals who filled in an online consent form. During the data collection process, 2 participants' data were removed from the study due to very poor audio quality of recordings, and 2 new interview slots were created to compensate for this.

The final participant sample consisted of 15 drivers who ranged in age from 25 to 64 (mean age = 45) and reported their gender as male (n = 8), female (n = 6) and non-binary (n = 1). The car brands represented within this study were as follows: Tesla (n = 6), Volvo (n = 3), BMW (n = 2), Jaguar (n = 1), Kia (n = 1), Land Rover (n = 1).

4.2. Materials

Materials included an online sign-up form, which was hosted on the Qualtrics platform (https://www.qualtrics.com) and consisted of: a Participant Information Sheet, consent form compliant with the General Data Protection Regulation 2016 and the Data Protection Act 2018, a set of questions related to car type and car use, a demographics section and a request to specify interview availability.

Materials also included an indicative interview script. The interview script was created by A1, with the aim of covering topics related to connectivity, data collection, data use and data privacy in cars. Creation of the interview script was guided by the principles of the HDI framework:

Legibility - to cover topics related to *Legibility* in the connected car context, we drafted 7 questions, focused on drivers' knowledge about, and attitudes towards, collection of data via their cars. Questions included, for example, *'If you compare your car and your phone – do you feel well informed about the data your car collects, in comparison to the data your phone collects?*'.

Agency - to cover topics related to *Agency* in the connected car context, we drafted 6 questions, focused on drivers' subjective and desired sense of control over the data collected by their car. Questions included, for example, *'Specifically in relation to your own car - do you feel that you have control over the types of data your car collects and the types of data it doesn't collect?*'.

Negotiability - to cover topics related to *Negotiability* in the connected car context, we drafted 7 questions, focused on drivers' opinions about who should create and/or manage car privacy norms, as well as 5 main questions focused on how/whether drivers' attitudes to car privacy may change across time and contexts. Questions included, for example, '*If car companies started suddenly collecting a lot more information about drivers and passengers inside the car, would that influence how you use your car?*'.

4.3. Procedure

The advertised link allowed participants to access the online survey, hosted on the Qualtrics platform. Following the Participant Information Sheet and consent form, participants were asked to input their contact information, and fill in a series of questions related to car type and car use, as well as a series of demographics questions. Participants were then asked to indicate their interview availability (at first this was only provided in the form of a multiple-choice question, with an option added during the data collection process to input more specific information about individual availability).

Video-recorded interviews were conducted by A1 and A2 via Microsoft Teams. The interviews lasted on average between 30 and 60 min. The researchers followed the indicative interview script; however, they were able to ask follow-up questions to explore topics further or to skip questions if the participant had already weighed in on a particular topic or when the assigned time slot was nearing the end. As such, there was variability in interview length, however, for most participants, all key topics were covered.

Following the interviews, the participants were informed about their right to withdraw from the study, received verbal debriefing, and were able to ask questions. Participants received £30 Amazon vouchers via email.

4.4. Data analysis

Data analysis was conducted by A1, employing the reflexive thematic analysis approach (Braun & Clarke, 2014). Interview recordings were transcribed using automatic transcription software. While checking transcripts for accuracy, the researcher highlighted sections pertaining to the research question. The transcripts were exported into Microsoft Word files, and the highlighted sections were subjected to open coding. The researcher then identified initial themes and reviewed the whole transcripts against them; this review process involved iterative recoding, where necessary, to improve consistency. The themes were then refined and named - which involved a back-and-forth comparison of codes and themes, reorganisation of data extracts across themes, and creation of written descriptions of themes, until coherent themes were arrived at. Finally, the researcher engaged in repeated reading of all data extracts assigned to each theme, while writing the Results section. The Results section was refined by A1 and A2 to ensure clarity.

We identified a total of 5 themes:

Theme 1: The mobile phone is a gatekeeper.
Theme 2: Drivers are both worried and hopeful about the outcomes of data collection.
Theme 3: The cabin is a private and personal space.
Theme 4: Road safety is a priority.
Theme 5: Limited knowledge about data collection means limited control over data.

5. Results

5.1. The mobile phone is a gatekeeper

This theme describes the ways in which mobile phones enabled drivers to experience connectivity and data collection in their cars. Mobile phones also influenced drivers' privacy attitudes in the car context.

5.1.1. Using the phone to make the most of the car

As participants used connected car applications on their mobile phones, they learnt about how their cars worked and what data were collected. For example, one participant realised that their car had connectivity and collected and stored data because they had to pair it with their phone: *'I feel like you connect it because you've got, I had to, like, pair my phone to the car. So it's stored that, it did ask me if I wanted to connect my contacts list to the car.*' **P1** For participants with limited interest in car connectivity, the mobile phone was a gatekeeper of some of the car's functionality. The simple act of connecting the car with the phone made them aware that the car was capable of connectivity: *'I mostly just use it for my, connecting it to my phone. So I can either play music or have, um, the satnav on the, uh, on the speaker in the car.*' **P11**

Several participants used proprietary mobile applications from their car companies. These applications facilitated communication with the car company. One participant felt that the app was the fastest way to obtain a response. They said: *'My most effective method of communicating with Tesla, in that I get the quicker response, is to ask them a question from within the app.*' **P9**. Another participant acknowledged the hybrid experience of using the car and the phone, stating that they were *'probably better informed through my phone about my car than from the car itself.*' **P15**.

Using a proprietary app helped drivers find information about their car. One participant bought a second-hand car and did not receive a manual. Using the app helped them gain access to one: '*I had to do a lot of digging on how to actually find the manual specifically for my car, for the for the type that I have. So, I had to download the BMW app and then from there, try to* log *in into that and access the manual for my car*.' **P11**. For some drivers, proprietary applications from their car company were either the sole or the primary means of accessing the data collected by their car: *'You know, the only data I would know how to access is what's what's there in the app.*' **P4**. Participants who did not have a hobbyist interest in car connectivity tended to assume that if some aspect of data collection was not in the proprietary app, then it was not taking place.

Participants with better awareness of car connectivity sometimes used these proprietary apps from their car company together with additional sources of mobility data. A Kia driver enjoyed using the external Ohme charging app to schedule their charging, despite this feature being available within the proprietary Kia app. There were, however, cases, where participants could not access the data that they wished to view via the proprietary app. This led drivers to seek out their own solutions. For example, some Tesla drivers used a third-party website called TeslaFi, as it provided more comprehensive access to car data, in a user-friendly format. One driver said that TeslaFi offers *'basically all your data in a way that you can read it.*' **P9**. This illustrates how third-party websites helped drivers both access and understand information collected by their cars, where this information was not made readily available by car companies. However, for this website (TeslaFi) to access the data collected by the car, drivers handed over access to their car's sensitive details (car account username and password), which made them concerned about cybersecurity. Insufficient access to data via the proprietary app was therefore ameliorated, but with what was seen to be a cybersecurity risk.

5.1.2. Access to proprietary app equals feeling of control over the car

An interesting example of the mobile app acting as gatekeeper, was the case of a participant who drove a Tesla provided by their employer. Despite being able to access the data in real-time on the dashboard, they were conscious of the fact that they did not have ownership of, and lacked the ability to oversee, the data about their journeys: *'I don't have the app for the car, but a manager at work has the app. And you do sometimes think, I wonder what's in that app, like, does it show how fast I've accelerated the car.*' **P4** Here, the disconnect between data subject and data oversight led to concerns about what inferences could be drawn from driving data. Moreover, it is notable that this user did not have access to the data merely because they were not the owner of the car, despite being the data subject.

5.1.3. Mobile phone as a point of privacy reference

Drivers spoke of data privacy in their cars by making references to their mobile phones. The breadth of data that mobile phones collect was seen to put car privacy issues in perspective: *'All of us carry a smartphone. And I figure that the smartphone already knows far more about me. So, the car, you know, it's just a very small picture of my usage of, you know, of where I go and who I'm with.*' **P3** When we initially spoke to participants about data collection in cars, they saw cars' ability to collect data as limited in comparison to mobile phones. One participant summed this up by saying: '*The last thing I'm worried about is the car logging where the car is being. Because my phone knows, you know, not only where I've been, but once I parked up my phone knows where I went when I got out of the car. And, you know, it's a lot more personal, that level of insight in my data*.' **P4**. Here, we see a potential risk to either not explore or discount one's own privacy preferences in relation to car data collection – simply due to the existing prevalence of location tracking in our daily lives.

5.2. Drivers are both worried and hopeful about the outcomes of data collection

This theme describes how concern about data privacy was less a general, binary, or even measurable concept but rather a complex web of expectations, assumptions, worries and preferences about how driving data may or may not be used.

5.2.1. Tension around how data sharing affects fairness and justice

Participants worried about whether the outcomes of data collection, processing, and use, would be fair and just. This generated arguments both for and against data collection. For example, one driver supported data collection to train machine learning models to help create autonomous vehicles. While this participant recognised the right of other drivers to refuse their data being utilised for this purpose, they believed that drivers who do refuse ought not to reap its rewards: *'I think it's a fair trade off to say, well, you can't then benefit from everyone else's driving data and use the self-driving features in the car that depend on people sharing that data.*' **P4** Another participant worried about whether insurance would remain fairly priced if reckless drivers refused to share their driving data with insurance companies: *'Is it fair if everyone pays more insurance?'* **P9**.

In parallel, doubting fairness of outcomes led to concerns about how their driving data could be used. This was the case even for those who fostered an attitude of openness. For one participant, the prospect of car companies collecting more data, was not likely to change how they used their car: '*Because I don't really do anything special with the car. You know, it only, only just takes me from home to work and work from home*.' **P2** The same participant, however, did have concerns about how data could be used in the context of automated processes. They experienced a road near their home being marked with the incorrect speed limit (30mph rather than 60mph) and worried about not being able to question erroneous databases: *'You know you're not able to protest that data. You're not able to say, Hang on a minute. We've been on one journey in the last month. You tell us we were speeding. I was the driver on that journey. I can tell you that we didn't go above 60*.' **P2** While the driver knew that they were obeying the law, the error led to an automatic assumption that they were not. Here, database errors and the lack of an appeals process, were seen to turn a mechanism supporting fairness into one that precludes fairness.

Richness of data held by firms supported trust in fairness. One participant stated that if offered, they would accept an insurance discount in exchange for sharing more data with their car company (Tesla): *'They have really meaningful data for which I can be rewarded if I'm a good driver or punished if I'm not. That's entirely fair*.' **P10** In contrast, some participants worried that insurance companies might not be able to untangle the nuances of driver behaviour. The risk of insurers making incorrect inferences made drivers worry about the fairness of premiums determined by an analysis of driving data. One participant, worried that an insurance company might not be able to distinguish between someone running through a traffic light because they are ignoring the rules and running through a traffic light because it is not safe to break at that point. They said: '*There is always going to be circumstantial situations where you might not follow the exact same rules on the road.*' **P11**. Conversely, when the outcome was to correct an unfair assumption; participants supported this use of data for insurance purposes. One participant liked the idea of data-driven Tesla insurance (not yet available for UK customers) as, due to many insurance companies classifying Tesla as a performance vehicle, this driver paid a premium for a feature that was not relevant to how they used they car: '*I just want a vehicle that will go a long way. I have to pay that premium because because it's capable of going nought to 60 so quickly*' **P3**.

Similarly, the ability of a company to prove a driver's innocence supported acceptance of data collection and retention. As one driver stated: '*If I have an accident, the Volvo can grab the data from the car's computer to work out how fast I was going when the brakes came on, when the lights were turned on.*' **P5**. This participant liked the idea that, in the case of an accident that was not their fault, the car company could use stored driving data to advocate on the driver's behalf: '*You know, this data here says exactly what the car did.*' **P5**

5.2.2. Track the car, not the driver

Several participants worried about becoming targets of unwanted commercial attention. Such unwanted attention could take the form of profiling, cold selling, targeting, or influencing – especially in the context of third-party companies – as well as merely annoyance. In other words, drivers worried about being bothered when they want to be left alone – in the context of using a car, a context which they associated with freedom.

Drivers felt more comfortable about data pertaining to their car being collected – as opposed to data pertaining to them as individuals. Some were keen to avoid targeted advertising, with one participant saying: '*I'm happy, um, from a from a technical and technology perspective, for the right people to have it. But I'm not a, I'm not a, what you call it, a target for sales.*' **P5**. Others wanted to make sure that their behaviour was not submitted to undue oversight. As one participant emphasised: '*If your car is collecting lots of data and then that can be reported back and shared, you know, with who knows who, you know, that doesn't, that kind of is the opposite of what I feel about freedom.*' **P14**. While this driver was happy with the car's condition being monitored, they were not happy with their behaviour being monitored: '*Stuff that car collects about itself, I, you know, and I'm not so bothered about, but where it pertains to kind of where I go and what I'm doing in my behaviour, then, yeah, I feel a bit more uncomfortable about that.*' **P14**

Drivers enthusiastically supported data collection that aimed at improving the vehicle. As one participant said: '*I feel the car is collecting data to, well, try and make the car better.*' **P3**. This participant also noted how such data collection was different from mobile phone tracking that, in their view, primarily served the purpose of advertising. Data about the car's performance were seen as directly relevant to product improvement: '*The only benefit in sharing the data is performance data in the product, product development, so I'll get a better product, better car to drive if they get driving data back from me.*' **P6** This was an example of a mutually beneficial interaction between drivers and car companies. For some drivers, this focused on a specific aspect of product improvement that they were interested in: '*I think it's extremely helpful that Tesla know what the performance of my car is, the battery degradation, whatever it might be, that sort of technical data, I certainly want to continue sharing.*' **P7**.

Where data collection did not support product improvement, drivers emphasised the importance of having the right consent procedures in place. As one participant explained, using car performance data for product improvement was very different from sharing such data with other companies: '*That's how they develop their products. They see how the product is being used and they develop it accordingly. If they are then selling that data to the third party that says, how many times I went to Aldi* versus *how many times I went to Waitrose because where I parked in the car park and at what times of day I do my shopping, then that's something which I think they ought to have asked for explicit consent for in terms of sharing that data.*' **P6**

Lack of granular consent procedures was a problem for some drivers and precluded them from navigating the sometimes worrying and sometimes useful context of data collection. They instead had to rely on switching telematics on and off and in turn losing useful features. As one participant explained: '*If you want to see the state, the charge of your vehicle on the on the mobile app, you need to accept everything.*' **P3** Drivers also felt that they needed to be given more information about data collection, to help make the right choices for themselves: '*It could say, well, you can pick and choose, but you'll lose certain features if you deny the access. And here's why we need the access. So, explain to me. If I'm giving you access to my data, what do I get for it?*' **P6**.

5.2.3. Importance of anonymity

Drivers expected the emphasis to be on data supporting performance while allowing them to remain anonymous: '*The thing is, can it be anonymised and aggregated into a, into a dataset which isn't identifiable back to you as an individual? And then it's, you know, it's more about the safety and reliability of the vehicle and not about, uh, and not about you as a driver.*' **P14**.

Aggregation of data made location tracking acceptable. As one driver explained: '*They're going to be interested in knowing things, like how many miles per day on average, across our whole fleet, do our cars drive? And I'm totally fine with the information being used for that because it's obviously totally anonymised.*' **P3**

The idea that data would be aggregated and not linked to individuals also led drivers to greater acceptance of wider access to data within car companies. One driver said: '*I don't know that limiting it within the organisation is helpful because I'm assuming that the data would be anonymised, so it would not connect to me. There's no need for that data not to be anonymous.*' **P6** It was pointed out, however, that aggregation of data may not always equal anonymity and in such cases car companies must take steps to protect the user's identity: '*If you live in the Highlands of Scotland and you're the only person with a Tesla in a 10-mile range or a Volvo, um, you wouldn't want Volvo then saying, you know, an average driver in the Highlands of Scotland did this. When you're the only one there and it's obvious it's you. You know what I mean. They've got to keep it anonymous. It's important. And possibly even anonymised to the car companies as well. There's no reason that they even need to know who you are.*' **P4**.

The possibility of losing anonymity led to concerns about safety and personal freedoms. For example, some participants were worried about the potential of car companies sharing data with law enforcement for proactive surveillance. Others were concerned about personal safety: '*For me, it's a bit like the conversation that's being had about Strava at the moment and running apps. And does, do you share the, you know, who do you share the location with. And is it, is it, you know, and also is it a woman's problem?*' **P14**.

5.3. The cabin is a private and personal space

This theme describes how, even among participants who showed openness to data collection, there was strong opposition to the collection of audio data in the cabin. Moreover, for most participants, there was opposition to the collection of data via internal cameras. This stemmed from concerns about intrusion into what was seen as a personal and private space.

5.3.1. A distinction between data about the car and data about the people inside the car

Participants expected privacy inside the cabin. Collection of audio and video inside the car would violate that expectation. As one participant said: '*I would no longer feel that the car is my private, safe environment that is private. And it's mine. You know, I would feel as though actually, at any point there could be somebody effectively sat in the car with me listening to what I'm saying.*' **P4**. This would then affect their use of the car: '*It would definitely impact how I would feel about the car and use it.*' **P4** Another participant likened their car to a living space, where recording of video and audio would be an intrusion: '*The car, in a sense, it's a bit like your, like your living space when you're in it. You have, you know, you have personal conversations with people.*' **P3**. Recording of video and audio in the cabin was perceived as crossing a boundary – as it would mean collecting information specifically pertinent to the driver and to the passengers. As one participant said: '*That's really, um, overstepping the line in terms of privacy. Sort of spying on the driver and the other people in the car. It's a bit… I think that's pretty outrageous.*' **P13** For this participant, the recording of audio or video in the cabin would be the one privacy-related factor that could influence their consumer choices: '*I think the only thing that would influence me would be if I, if I found out that a car that I was looking at, um, could record, you know, whether that's visual or audible activity.*' **P13** For another driver, visual or audio surveillance would be enough to consider switching to a different car make: '*I think that would actually put me off having of the car. If the audio and video was being recorded. Um, I think I'd look for a different car*.' **P15**

5.3.2. The safety-privacy trade-off of internal cameras

The safety-supporting role of an internal camera was a divisive topic. For some, the potential safety benefits of such cameras outweighed privacy risks, even to the point of willingness to give up control over privacy choices to car companies: '*I would half like the option to turn off the cameras, but also when I put on my, uh, safety head, I don't want the option to turn off the cameras because they're providing useful features like helping with the collision avoidance and helping, say, if it's a long trip, you're falling asleep*.' **P9** This participant told us that, for them, the safety benefit outweighs the loss of confidentiality.

Others felt that the safety benefits were merely potential and did not outweigh the existing and real privacy risks: '*The chances of you falling asleep at the wheel compared to the amount of other stuff that they would be peripherally collecting, um, I can't see what the benefits to me would be of a camera inside the car, really*.' **P14**

These two opposing views could potentially be reconciled: another participant said that the collection of internal video data could be acceptable, if it remained contained within a closed loop system: '*Maybe it's used for the onboard computer only, and the data isn't stored or recorded. It's just used as a sensor. Am I looking at the road? Yes, I am okay. No, I'm not looking at the road anymore, so the car beeps and warns me, but it's not sending that data off to anybody else*.' **P4**

5.3.3. Opposition to collection of internal audio data

Opposition to the collection of audio data, from inside the cabin, was universal. Drivers wanted to be able to speak freely and not worry about what information could be picked up. One driver explained: '*Because there are conversations that I might have with my, with my family or friends, that maybe have sensitive information in them. Personal information about us and people that we know*.' **P12** This participant also emphasised the importance of your vehicle being a personal space where you can have private conversations.

Concern about the recording of personal conversations was present even for drivers who were not otherwise privacy conscious. For one participant this was caused by existing suspicion about the collection of audio data in other contexts: '*If the car is collecting audio data, that's something I certainly would want turning off. And it's something that worries me about the phone because I do have suspicions that Facebook have been using audio data*.' **P7**

If recording of conversations could not be avoided, this would potentially influence how drivers would behave inside the cabin and how they would use their cars. One participant said, if audio were to be recorded: '*I would just try not to have any kind of sensitive conversations whilst driving*.' **P11**

5.4. Road safety is a priority

This theme describes how prioritising road safety attenuated our participants' privacy concerns. Road safety emerged as an overarching value that changed the drivers' priorities, for participants who were otherwise particularly privacy conscious. Nevertheless, prioritising road safety did not mean that drivers wished to give up all controls over their car data.

5.4.1. Trading privacy for the sake of road safety

Drivers supported data collection, if it facilitated road safety, for themselves or for other drivers. This could take the form of real-time assessment of road conditions: '*One of the things that I like about the Volvo, the Volvo system, is that it collects, um, connected safety data. So, for example, if another car in the network experiences like a wheel slip or has harsh braking, it sends it to Volvo Service, and the Volvo tells all the cars there are in an area that there's problems. And that thing, I think, is quite useful*.' **P5**.

Another aspect of road safety was companies' ability to monitor performance to identify potential problems early on: '*I mean, there has been a recall on my car because of problems with, you know, reported problems with models. And I guess if you can speed that up*.' **P14**.

5.4.2. Tensions between road safety and privacy

Some participants were willing to trade data privacy for road safety. However, this should not be misinterpreted as not wanting any control at all. Drivers wanted to get help in the case of an accident - which created acceptance of real-time location monitoring: '*I haven't really got any problems with them knowing where I am because I've read before, I don't know how true it is, that if you have a serious accident in the car, it alerts Tesla and they can send a technician out to check your car over and make sure it's safe to drive*.' **P8**. However, participants wanted such data use to remain within the boundaries of this specific purpose: '*But what I'm sort of not comfortable with is them being able to interrogate where I've been over a period of time and draw conclusions about me*.' **P14**

5.5. Limited knowledge about data collection means limited control over data

This theme describes the ways in which drivers were aware of limits to their knowledge about what data are collected as they drive their cars, and what happens to these data thereafter. Some participants saw their limited understanding of the data life cycle as a barrier to having control over their data.

5.5.1. Causes of limited knowledge about data collection

For some participants, limited knowledge about the data their car collected resulted from lack of previous interest. One driver explained that their car '*could collect data using the satnav, the in-car system, which then when I have it serviced at the dealers and it gets a software upgrade, could be then translated, you know, could be transferred back to the, uh, to Volvo. But I don't know, I haven't thought about it in great depth*.' **P14** Notably, for some participants the interviews we conducted were the first time that they considered what data their cars might be collecting. One driver said: '*I would be interested to know what data they are collecting, actually collecting on the car, just out of interest really. Um, it's not something I've ever asked of Tesla*.' **P7**

One of the interesting aspects of this research was seeing the participants' awareness of and thinking about data collection in cars, develop. Most participants simply had not considered the concept of data collection in cars yet. Most participants initially assumed that data collection would be of more relevance to car companies, or that the information about these issues might not be accessible or easy to understand. Once participants started to think more deeply about data collection and use, questions emerged.

Participants felt that car companies could do more to inform them on the topic of data collection. As one participant said: '*There doesn't seem to be that, um, responsiveness, uh, on a customer service basis. I think the the terms and conditions of data protection and customer consumer protection tend to be a little bit officious. And actually, there should just be a help line*.' **P2** Participants also felt that, when it came to informing drivers about data collection, car companies were not trying to communicate effectively: '*I did go looking and apart from reading lots of user licence agreements for lots of different systems in the car… But they don't really talk about what they're collecting and why*.' **P5**. This contrasted to car companies having effective communication strategies for marketing and maintenance: '*I don't think they are proactive in communicating their policies. They are proactive about communicating with me about sales of new vehicles and when my car needs servicing and that kind of thing, but less so about any data collection*.' **P14**

5.5.2. Consequences of limited knowledge about data collection

Limited availability of information about data flows created scepticism about the degree to which car companies were transparent in what they did or did not collect. One of the participants who was knowledgeable about data privacy and had a hobbyist level interest in car connectivity said: '*I'm certain that the API doesn't doesn't give you all the information they're collecting. They will be collecting much more than that. Um, so in, in terms of, if you wanted an exhaustive list of everything, I don't think that's available other than going to these, you know, seeing what this kind of thing that these security researchers and reverse engineers [study].*' **P3** This demonstrates that, even for one of our most knowledgeable participants, understanding data flows in their car was a challenge.

Advanced knowledge about data privacy and the systems involved in car connectivity did not protect drivers from frustration. One participant shared that although they were happy with what the company currently communicated, they were also aware of a planned move to a different in-car third-party operating system: '*It's just they don't make a very good job of, now they're moving to a new platform, of what actually are privacy implications.*' **P5** Here, we see how understanding of the data life cycle in the context of one's car requires updates and consistent communication from the company to the user.

5.5.3. Consent: Amount, context, and lifecycle of data (It's overwhelming)

Limits to knowledge about the data life cycle challenged drivers' ability to exercise ownership over these data. One driver felt that they did not have much control as they had not investigated what data are collected: *'It's probably collecting all sorts of data right now and I've got no clue.*' **P1** Another participant started to question whether simply using a connectivity feature (GPS) meant that they were unwittingly consenting to data collection: *'It could be collecting data about the times of day I'm using the car or my journeys, my charging patterns and and history. I don't know. I don't, I guess I'm saying I don't know what it's collecting, but I guess it has the capacity to collect those things. And maybe by having it activated, I'm implicitly giving, giving it permission to do so.*' **P12**. This quote showcases how, when deprived of information about what data their cars collect, drivers cannot participate in informed consent.

6. Discussion

Our data analysis identified 5 overarching themes. We found that: (1) the mobile phone plays a key role in helping drivers understand their car's connectivity features, (2) drivers are both worried and hopeful about what the use of their car data may lead to and how it might impact them personally, (3) drivers expect the cabin of their car to remain a private space where their behaviour and conversations are not monitored, (4) road safety is a priority for drivers and safety concerns may result in more lenient attitudes to data privacy in cars, and (5) when drivers do not receive adequate information about what data are collected via their cars, this limits their ability to feel in control.

Below, we discuss what the three tenets of the Human-Data Interaction network mean, based on our findings, in the context of the connected car. This is followed by design recommendations – for professionals designing connected car technology – for supporting the legibility, agency and negotiability of automotive data. Finally, we discuss the implications of this work in the context of existing research.

6.1. Meaning of Legibility, Agency and Negotiability in the context of a connected car

Within the Human-Data Interaction framework, 'Legibility' is 'concerned with making data and analytics algorithms both transparent and comprehensible to the people the data and processing concerns' (p.4) (Mortier et al., 2014). In the context of a connected car, these people would include drivers, passengers and passers-by, however this paper focuses specifically on drivers. Based on our findings, we believe that good Legibility of automotive data means that:

*Drivers know what data are collected through their cars, how these data are stored and how they are used. They understand the connections between data collection and their ability to make the most of their cars (*e.g., *in relation to tracking battery usage or connectivity features such as GPS). They also know about any secondary uses of car data such as for marketing or resale to third parties. Drivers feel confident about accessing and reviewing the data gathered by their cars. Finally, drivers feel confident accessing information about car data protection and car data lifecycle from the car manufacturers.*

Within the Human-Data Interaction framework, 'Agency' is 'concerned with giving people the capacity to act within these data systems, to opt-in or to opt-out, to control, inform and correct data and inferences, and so on' (p.4) (Mortier et al., 2014). Based on our findings, we provide the following summary of what good *Agency* of automotive data means in the context of connected car drivers:

Drivers can easily access data-sharing preferences both in the mobile app and on the car's dashboard. These systems are designed to promote decision-making based on how drivers interact with car data, for example by linking data-sharing settings to information about outcomes of data collection and by seeking separate and distinct consent for data that drivers are particularly concerned about, that is internal cabin audio and video.

The Human-Data Interaction theory defines 'Negotiability' as 'concerned with the many dynamic relationships that arise around data and data processing' (p.4) and emphasis is placed on providing 'support for people to re-evaluate their decisions as contexts change' both externally and internally (p.7) (Mortier et al., 2014). Based on our findings, we provide the following summary of what good *Negotiability* of data means in the context of connected car drivers:

Drivers can refine and change their data-sharing preferences based on changing circumstances. Changes in circumstances are proactively communicated by car companies. The option to adjust data-sharing settings is available continuously. Settings are sufficiently granular to support changes that arise from new circumstances, and easily accessible within the mobile app and on the dashboard.

Below, we outline the design recommendations created as a result of our findings, to facilitate Legibility, Agency and Negotiability in the context of car data. As the current study is an exploration of drivers' needs and not an assessment of current industry practices, our recommendations are universal and do not relate to, nor comment on, specific systems currently in place.

6.2. Design recommendations to support Legibility of car data

6.2.1. Present car data, and information about its use, within a mobile app

Car companies should provide a proprietary app, with clearly signalled features that help drivers understand what data are collected via their cars. In this way, a mobile app should serve as a means of communication between the driver and the connected car. The familiarity and ease with which drivers interact with their phones can help make the complex world of car data more accessible.

6.2.2. Inform drivers about all outcomes of data collection

Our findings indicate that drivers care about how collection and use of data translates into outcomes that are useful for them and for other drivers. Car companies should communicate about how collection and use of data serves drivers. Simultaneously, information should be given about outcomes of data collection that drivers might be less keen on, for example surveillance of individuals, or third-party advertising. Our findings suggest that understanding the outcomes of data collection is a central aspect of car data legibility.

6.2.3. Promote/push information about audio and video data inside the cabin

Our findings indicate that the collection and use of audio and video inside the cabin has particularly meaningful consequences for whether drivers feel that their privacy is being respected. Therefore, achieving an understanding of what data are collected inside the cabin and what happens to these data later is crucial for achieving legibility of car data. Drivers should be provided with clear, complete and up-to-date information about sensors inside the cabin and the data they collect, as well as the justification and uses for such data collection. Communication about cabin surveillance should involve regular updates and notifications that make it easy for drivers to understand how changes in cabin surveillance affect them. Considering the emotive responses of our participants in conversations about audio and video collection inside the cabin, we suggest that providing information about in-cabin data collection via both app and dashboard should be a top priority.

6.2.4. Explain the links between data collection and road safety

Our participants were willing to embrace data collection, should it result in better road safety for themselves or others. As part of promoting legibility of car data, car manufacturers should highlight links between data collection and road safety. This could be through regular updates via the dashboard and the mobile app.

6.2.5. Facilitate proactive and responsive communication

In our study, drivers were sceptical about companies' transparency. They often assumed that lack of legibility was intentional and served the purpose of obfuscation. Car manufacturers that wish to support a trusting relationship between themselves and drivers need to prioritise legibility and provision of information about what data are collected and what happens to these data. Companies that fail to inform users about aims and processes of data collection may risk giving the appearance of secrecy. Mobile apps and the dashboard should proactively encourage the driver to learn about data flows and the consequences of sharing or withholding specific types of data. For some drivers, the ability to only interact with a digital tool may be insufficient to create a sense of transparency and therefore being able to talk to someone should be incorporated into the communication – this could mean the ability to call a helpline directly from the app.

6.3. Design recommendations to support the Agency of drivers

6.3.1. Allow data-sharing decisions to be made via the mobile app

Designing a proprietary app to provide a full overview of what data are collected and then allowing drivers to opt in or out of certain types of data collection could improve user agency. It would be especially impactful for those with less knowledge about car connectivity, or for drivers who find the phone app more intuitive than using the dashboard. Another important way of supporting agency through design would be to improve the access to existing driving data, as some drivers may wish to monitor their car use and their car's condition or may want to see historical data on their phone.

6.3.2. Allow drivers to view and challenge outcomes of data collection

Drivers should be able to make granular data-sharing choices based on how they wish to use their car, and which types of data collection they support. Data-sharing settings should provide the information and granularity needed to support this. This relates to direct outcomes such as product improvement as well as more nuanced ones such as losing anonymity.

Moreover, our participants worried that if important decisions affecting them were made based on the driving data in a way that is automated, such decisions would be hard to challenge. Providing customer support in this area could help support drivers' sense of agency.

6.3.3. Make data-sharing controls for audio and video inside the cabin separate

The agency of users is particularly important in the context of the inside of the cabin. Drivers see the car cabin as a private space and so companies need to pay special attention to providing consent procedures for whether and how data from the cabin are collected even when these data may support safety or innovation. These data-sharing controls should be distinct and separate so that users may be certain that they are in charge of audio and video surveillance inside the cabin.

6.3.4. Where it is legal and practical, allow users to balance safety and privacy according to their values

While allowing users to switch off safety features to disable data collection would be neither reasonable nor, in many cases, compliant with the law, cars with advanced connectivity may be able to offer users some degree of choice when it comes to facilitating safety. For example, some users may wish to continuously transmit their location so that they may be tracked in the case of an accident, while others may choose to opt out of this feature. Car manufacturers should create safety-specific data-sharing controls, linked to clear explanation of how a certain type of data collection supports safety, so that users can balance safety and privacy.

6.3.5. Provide separate controls for primary and secondary uses of data

We found that drivers, not yet well versed in what data are collected and used by car companies, focused on the connectivity features these car data enabled (e.g., GPS, infotainment). This could lead to the assumption that no data use would happen, beyond facilitating the features of the car that drivers themselves benefit from. This signals that drivers need support around managing their data-sharing preferences. Such concerns must be addressed when designing data-sharing settings. To support drivers' Agency, car manufacturers should design data-sharing settings in a way that links information about data collection with information about how data are used. It should not be assumed that drivers consent to secondary uses of data collected to facilitate a connectivity feature (e.g., location data in the case of GPS). Data-sharing settings should be designed to separately seek consent for primary (to enable a feature) and secondary uses of car data.

6.4. Design recommendations to support Negotiability of car data

6.4.1. Allow smooth transition between app and dashboard to enable context-appropriate data-sharing decisions

Based on personal preferences and experience, drivers may choose to make data-sharing decisions on the dashboard, within the app, or in both these settings. In our study, some users, typically with a hobbyist interest in connected cars spoke predominantly of the dashboard. Drivers without much knowledge about car connectivity referred more often to the mobile app. The drivers' needs may change over time. A driver that once relied on the app may became more comfortable using the dashboard. To support personal preferences and the evolving expertise of drivers, designers should create data-sharing settings that can be accessed from both the mobile app and the dashboard. The look and feel of such settings should also be aligned so that drivers can seamlessly move between the two.

6.4.2. Regularly communicate outcomes of data collection and allow users to adjust data-sharing settings accordingly

We found that drivers often base their data-sharing preferences on how the data will be utilised and whether the outcome might be useful to them or harm them personally. Drivers need to be continuously informed about how car data are used – both from the perspective of what the company aims to accomplish (e.g., improving the car versus monetisation of data by sale to third parties) and in terms of what inferences are drawn from the data and how these may affect individual drivers. Moreover, drivers care about aggregation and anonymity, and they should also be continuously informed about how their data are handled – which data are anonymised, and which can be linked to an individual, and by whom. Embedding this information into the data-sharing settings will allow drivers to revise their data-sharing preferences based on the changing context of the consequences of agreeing to share one's data with the car manufacturer.

6.4.3. Support users in managing data-sharing preferences for in-cabin audio and video

In our study, drivers were highly concerned about audio and video data from their cars being collected, and were particularly worried about the idea of their conversations being listened to. At a high level, car companies have two options. Firstly, they may need to decide not to collect this type of data, accepting that for many drivers this will be a privacy boundary. Secondly, it may be possible to still make use of various types of data from the cabin so long as the analysis and the use of these data are designed within a closed loop system. For example, a car could record and interpret voice commands in the cabin without those data leaving the car. This could mean that connected cars may need more processing power to become more self-sufficient within a closed loop.

Where companies decide to collect in-cabin audio and video it may be crucial to involve drivers in decision-making. In our study, drivers wanted to retain a sense that the cabin is a private space in which they can feel safe to have private conversations. Negotiability, supported by drivers being able to turn non-safety critical sensors on and off may be particularly important in the case of surveillance inside the cabin.

6.4.4. Enable adjustment of data-sharing preferences based on how collection affects road safety (across routes and weather conditions)

It is possible that, in some circumstances, extended data collection may support the safer use of a car, but in order to do so, has to cross some of the privacy boundaries and preferences drivers have. Drivers may wish to make independent trade-offs (where legally and practically reasonable) between safety and privacy. For example, some drivers may prioritise road safety above privacy in all circumstances. Others, however, may be comfortable with providing some types of data (e.g., performance data) but not internal video. Others yet may lower their privacy expectations for the sake of safety in extreme weather or when travelling with children but may wish to return to more conservative privacy settings in other contexts. In sum, the decisions that drivers may wish to make about data-sharing and their ability to prioritise privacy versus road safety (where the two may conflict) should be supported across contexts. To do this, drivers should be regularly updated about how data collection impacts road safety and should be given opportunities to adjust data-sharing preferences following each update.

6.4.5. Prompt users to assess data-sharing preferences following major changes to data life cycle

As car manufacturers develop their business models and refine their approaches to the use of car data, drivers need to be kept informed about these changes, and how they affect them. For example, if a company employs a new data processor, makes changes to the software used for data-gathering, or decides to archive data for a shorter or longer period, these changes may have consequences for individual data-sharing preferences. Such information should be communicated not only within privacy policies and emails to customers but also directly next to data-sharing settings (for example, in pop-up boxes). This way, drivers can choose to adjust their preferences in line with the new context of data use.

Drivers should receive a clear explanation of what changes (to data processing or use) are taking place and how it will affect them, now and in the future. Data-sharing settings could include time-sensitive controls, to allow drivers to pre-plan their data-sharing, for example consenting to share a specific type of data until the next privacy update.

6.5. General discussion: Value-driven privacy preferences

The drivers we interviewed were primarily concerned with the outcomes of data collection, rather than being opposed to or accepting of the collection of particular types of data. Although not against location tracking, drivers did not want location data to be used to infer information about them. Research demonstrates that even seemingly benign location data can allow inference of personal attributes (Baron & Musolesi, 2020). While most of our participants were not necessarily aware of such implications, they nevertheless made a clear distinction between using car data to learn something about the car (which in turn supported the value of usefulness) and using car data to learn something about the driver (which went against the value of anonymity).

Furthermore, our participants were concerned about secondary uses of data – for example drivers emphasised that re-use of car data for targeted marketing would necessitate additional consent. This is in line with research demonstrating that consumers do not relinquish interest in what happens to their data following initial disclosure (Souza & Phelps, 2009). Drivers in our study had a clear idea of how data ought to and ought not to be used. They expected these ideas, and the values that guided them, to be upheld throughout the data life cycle. For example, a commonly held value was ‘usefulness’, relating to improved functioning of the car, now or in the future. Data uses aligned with this value were seen as legitimate. Data use that did not align with it was seen as requiring additional and more granular consent. This is in line with the contextual integrity theory of privacy (Nissenbaum, 2004). Our participants, even those without extensive knowledge about car data, had defined personal opinions about what was reasonable, and what was not, in the context of car privacy.

Interestingly, our findings are also, to some extent, in line with the privacy calculus theory, which assumes that users make data sharing decisions based on a calculation of risks and benefits (Kokolakis, 2017). For example, some drivers in our study prioritised road safety to such an extent that they were willing to trade privacy for it. Benefiting from improved road safety cannot, however, be compared to sharing data in return for a discounted service (a scenario common to ‘privacy calculus’ research) (Kokolakis, 2017). Instead, it was a deeply shared value that was context-appropriate (it is uncontroversial that both drivers and car manufacturers desire to support road safety) and did not involve secondary and unrelated uses of data.

Some values appeared quite universal across our participants. Consistent with the contextual integrity theory, drivers expected privacy inside the cabin so collecting audio and video in that environment was seen as particularly problematic. This is in line with the view that a car is a private sphere, associated with freedom and independence (Walter & Abendroth, 2018). Opposition to the collection of cabin audio was universal, suggesting that some norms in the car context may need to be adopted across the board. At the same time, the value of road safety was a mediator for how drivers perceived the collection of cabin video. Some drivers were willing to ‘trade’ privacy for ‘safety’ and allow in-cabin video surveillance. This occurrence is reflective of the privacy calculus perspective, with the caveat that most users will still expect additional consent for secondary uses (e.g., re-selling) of data. This suggests that car manufacturers may benefit from business models that maximise the use of data within the closed environment of the company (e.g., by benefiting from data-led innovation) and do not rely on re-sale. Our findings indicate that the former may be met with acceptance, or even enthusiasm from drivers, while the latter will be subject to far greater scrutiny from them.

Furthermore, this study demonstrates that while the contextual integrity framework can explain some of the privacy preferences in the context of car data, drivers may have differing values that determine and mediate their data-sharing preferences and so creating uniform privacy boundaries for all users may not be the best way of facilitating optimal privacy. Depending on prioritised values, drivers may wish to minimise data-sharing in some areas and maximise it in others.

It is possible that the contextual integrity lens and privacy calculus lens are complimentary in understanding how to best facilitate user privacy. Indeed, our findings suggest that employing the Human-Data Interaction framework to guide the enquiry can help researchers achieve a broader understanding of users’ privacy preferences and boundaries, reflective of both the norms that they expect while interacting with a particular technology as well as their personal priorities and the trade-offs that they are happy to make.

7. Limitations and future steps

The extent to which findings from this study can be seen as representative of the wider population of connected car drivers is limited by several factors. As the key aim of our investigation was to scope drivers' experiences and opinions - and as this is still an emerging and rapidly changing area - we did not delve deeply into *how* our participants' views on privacy and/or their data-sharing preferences do or may influence their *consumer behaviour*. Further research could help untangle both the impact of car data-sharing preferences on consumer behaviour, as well as the legal protections that drivers would like to see in place. Such research could help car companies and lawmakers navigate the areas of automotive data use, protection, and monetisation.

Furthermore, although the Results section evolved from a thorough and iterative examination of the interview data, we did not subsequently contact the participants to ensure that we had accurately interpreted their statements within the interviews. Future research focused on delivering solutions for better privacy and human-data interaction in the connected car context might benefit from engaging with a sample of participants across several stages of investigation. For example, it could be beneficial to design consent solutions for connected cars based on interviews and then conduct a trial of this technology with the same group of participants. This would allow the researchers to ensure that interviewees' privacy preferences are well understood and well implemented. Such research would help support the accuracy and validity of driver interview data in the context of connected car technology design recommendations.

Finally, our findings demonstrate how the Human-Data Interaction framework can be employed to navigate the complex web of privacy needs and expectations in areas where stakeholders may have reasons to *want* to share their data for the public good. It could be useful to apply this framework when developing climate-smart solutions for cities, for example by exploring citizens' willingness to share transport data, how this may change across contexts and what consent procedures are needed to ensure a sustainable flow of information.

8. Conclusion

We presented findings from interviews with 15 drivers of connected cars. Guided by the Human-Data Interaction framework, we developed design recommendations to support drivers in achieving better understanding of and control over their car data, across different contexts. This work contributes to the field of Human-Computer Interaction by providing a foundation for a user-centric study of data privacy in modern cars, by demonstrating how drivers make sense of data, what their priorities are and what they expect from car manufacturers with respect to data protection and privacy.

Funding

First author and last author received funding from *Human Data Interaction: Legibility, Agency, Negotiability,* a 'network plus' funded by the UK Engineering and Physical Sciences Research Council.

CRediT authorship contribution statement

Anna Dowthwaite: Conceptualization, Formal analysis, Investigation, Funding acquisition, Supervision, Writing – review & editing. **Dave Cook:** Formal analysis, Investigation, Writing – review & editing. **Anna L. Cox:** Conceptualization, Funding acquisition, Investigation, Supervision, Writing – review & editing.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Data availability

The data that has been used is confidential.

References

Baron, B., Musolesi, M., 2020. Interpretable machine learning for privacy-preserving pervasive systems. IEEE Pervasive Comput. 19 (1), 73–82.

Bertoncello, M., Martens, C., Möller, T., Schneiderbauer, T., 2021. Unlocking the full life-cycle value from connected-car data. McKinsey.

Braun, V., Clarke, V., 2014. What can "thematic analysis" offer health and wellbeing researchers? Int. J. Qual. Stud. Health Well Being 9 (1), 26152.

Buck, C., Reith, R., 2020. Privacy on the road? Evaluating German consumers' intention to use connected cars. Int. J. Automot. Technol. Manag. 20 (3), 297–318.

Caltrider, J., Rykov, M., & MacDonald, Z. (2023, September 6). After Researching Cars and Privacy, Here's What Keeps Us up at Night. https://foundation.mozilla.org/en/privacynotincluded/articles/after-researching-cars-and-privacy-heres-what-keeps-us-up-at-night/.

Derikx, S., De Reuver, M., Kroesen, M., 2016. Can privacy concerns for insurance of connected cars be compensated? Electron. Mark. 26, 73–81.

D'Souza, G., Phelps, J.E., 2009. The privacy paradox: The case of secondary disclosure. Rev. Mark. Sci. 7 (1), 0000102202154656161072.

Eski, Y., Schuilenburg, M., 2022. On Tesla: Balancing sustainable car connectivity, silent lethality and luxury surveillance. Criminological Encounters 5 (1), 234–251.

Fowler, G. A. (2019). What does your car know about you? We hacked a Chevy to find out. *The Washington Post*.

Frassinelli, D., Park, S., Nürnberger, S., 2020. I know where you parked last summer: Automated reverse engineering and privacy analysis of modern cars. In: 2020 IEEE Symposium on Security and Privacy (SP). IEEE, pp. 1401–1415.

Gurumurthy, K.M., Kockelman, K.M., 2020. Modeling Americans' autonomous vehicle preferences: A focus on dynamic ride-sharing, privacy & long-distance mode choices. Technol. Forecast. Soc. Chang. 150, 119792.

Hajlaoui, R., Moulahi, T., Guyennet, H., 2018. Vehicular ad hoc networks: From simulations to real-life scenarios. J. Fund. Appl. Sci. 10 (4S), 632–637.

Iachello, G., Hong, J., 2007. End-user privacy in human–computer interaction. Foundations and Trends®. Human-Computer Interaction 1 (1), 1–137.

Information and Privacy Commissioner of Ontario. (2016). De-identification Guidelines for Structured Data. Available at: https://www.ipc.on.ca/wp-content/uploads/2016/08/Deidentification-Guidelines-for- Structured-Data.pdf.

Jakobi, T., Alizadeh, F., Marburger, M., & Stevens, G. (2021). A Consumer Perspective on Privacy Risk Awareness of Connected Car Data Use. In *Proceedings of Mensch und Computer 2021* (pp. 294-302).

Jung, J.C., Sharon, E., 2019. The Volkswagen emissions scandal and its aftermath. Glob. Bus. Organ. Excell. 38 (4), 6–15.

Kaur, K., Rampersad, G., 2018. Trust in driverless cars: Investigating key factors influencing the adoption of driverless cars. J. Eng. Tech. Manage. 48, 87–96.

Kokolakis, S., 2017. Privacy attitudes and privacy behaviour: A review of current research on the privacy paradox phenomenon. Comput. Secur. 64, 122–134.

Lim, J., Kim, K., Yu, H., Lee, S.B., 2020. Making connected cars untraceable via dsrc radios. IEEE Access 8, 224932–224946.

Löbner, S., Tronnier, F., Pape, S., & Rannenberg, K. (2021, November). Comparison of de-identification techniques for privacy preserving data analysis in vehicular data sharing. In *Proceedings of the 5th ACM Computer Science in Cars Symposium* (pp. 1-11).

Lyon, D., 1993. An electronic panopticon? A sociological critique of surveillance theory. Sociol. Rev. 41 (4), 653–678.

Martin, K., 2020. Breaking the privacy paradox: the value of privacy and associated duty of firms. Bus. Ethics Q. 30 (1), 65–96.

Mortier, R., Haddadi, H., Henderson, T., McAuley, D., & Crowcroft, J. (2014). Human-data interaction: The human face of the data-driven society. *arXiv preprint arXiv: 1412.6159*.

Nissenbaum, H., 2004. Privacy as contextual integrity. Wash. l. Rev. 79, 119.

Rahal, R., Amara Korba, A., Ghoualmi-Zine, N., Challal, Y., Ghamri-Doudane, M.Y., 2022. AntibotV: A multilevel behaviour-based framework for botnets detection in vehicular networks. J. Netw. Syst. Manag. 30, 1–40.

Ram, P., Markkula, J., Friman, V., Raz, A., 2018. Security and privacy concerns in connected cars: A systematic mapping study. In: 2018 44th Euromicro Conference on Software Engineering and Advanced Applications (SEAA). IEEE, pp. 124–131.

Richards, N.M., Solove, D.J., 2010. Prosser's privacy law: A mixed legacy. Calif. l. Rev. 98, 1887.

Samarati, P., & Sweeney, L. (1998). Protecting privacy when disclosing information: k-anonymity and its enforcement through generalization and suppression.

Stephenson, S., Almansoori, M., Emami-Naeini, P., & Chatterjee, R. (2023). "It's the Equivalent of Feeling Like You're in Jail": Lessons from Firsthand and Secondhand Accounts of IoT-Enabled Intimate Partner Abuse. In *32nd USENIX Security Symposium (USENIX Security 23)*.

Victorelli, E.Z., Dos Reis, J.C., Hornung, H., Prado, A.B., 2020. Understanding human-data interaction: Literature review and recommendations for design. Int. J. Hum Comput Stud. 134, 13–32.

Walter, J., & Abendroth, B. (2018). Losing a private sphere? A glance on the user perspective on privacy in connected cars. *Advanced Microsystems for Automotive Applications 2017: Smart Systems Transforming the Automobile*, 237-247.

Walter, J., Abendroth, B., & Agarwal, N. (2017, November). PRICON: self-determined privacy in the connected car motivated by the privacy calculus model. In *Proceedings of the 16th International Conference on Mobile and Ubiquitous Multimedia* (pp. 421-427).

Warren, S., Brandeis, L., 1989. The right to privacy. In: Killing the Messenger: 100 Years of Media Criticism. Columbia University Press, pp. 1–21.

Zhang, J., Wu, M., 2021. Blockchain-based authentication with optional privacy preservation for internet of vehicles. Math. Probl. Eng. 2021, 1–13.

Zuboff, S., 2019. The Age of Surveillance Capitalism: The Fight for a Human Future at the New Frontier of Power. Public Affairs, New York.