Privacy in Transport? Exploring Perceptions of Location Privacy Through User Segmentation

Ingolf Becker[†]*, Rebecca Posner[‡]*, Tasmina Islam[§], Paul Ekblom[†], Hervé Borrion[†], Michael McGuire[¶], Shujun Li[‖]

[†]UCL, London, UK {i.becker,p.ekblom, h.borrion}@ucl.ac.uk

[‡]TRL rebecca.a.posner @gmail.com

[§]Kings College London, UK tasmina.islam @kcl.ac.uk

[¶]University of Surrey, UK m.mcguire @surrey.ac.uk

[‖]University of Kent, UK hooklee@gmail.com

Abstract

Unanticipated accumulation and dissemination of accurate location information flows is the latest iteration of the privacy debate. This mixed-methods research contributes a grounded understanding of risk perceptions, enablers and barriers to privacy preserving behaviour in a cyber-physical environment.

We conducted the first representative survey on internet privacy concerns, cyber and physical risk taking, privacy victimisation, usage of location sharing apps and transport choices in the UK with 466 participants. The responses segregated participants into four distinct, novel clusters (cyber risk takers, physical risk takers, transport innovators, and risk abstainers) with cross-validated prediction accuracy of 92%.

In the second part of the study, we qualitatively explored these clusters through 12 homogeneous focus groups with 6 participants each. The predominant themes of the groups matched their clusters with little overlap between the groups. The differences in risk perception and behaviours varied greatly between the clusters. Future transport systems, apps and websites that rely on location data therefore need a more personalised approach to information provision surrounding location sharing. Failing to recognise these differences could lead to reduced data sharing, riskier sharing behaviour or even total avoidance of new forms of technology in transport.

1. Introduction

Recent revelations about the use of Strava to identify American military bases in the Middle East [1] have revived the debate on location tracking. The motivations for collecting detailed, high accuracy location data are obvious: for public transport, automatic departure, disruption and routing information is valued. In connected and autonomous transport systems, location data sharing is attached to improvements in road safety, accident prevention, information for emergency and breakdown recovery situations, and other economic and environmental benefits for the driver, passengers and other road users [2]. Future transport systems will rely on an ever more interconnected, real-time information system that requires continuous location data from its users for reliable and efficient resource allocation. In addition to the privacy invasion of highly targeted advertising, there are also physical consequence stemming from the abuse of this data. Houses are often burgled when they are unoccupied [3], expensive bicycles are identified and located through social fitness networks [4], relationships and social activities can be monitored, and harassment and domestic abuse can be facilitated [5].

Previous research has studied the responses to privacy risks, and motivations and coping behaviours in great detail, as we will discuss in the literature review in Section 2. However, research on location privacy has exclusively focused on transparent, transactional privacy invasions, such as 'check-in' systems and location-based advertising systems where there is a clear, immediate reward for giving up one's location privacy for a short period. Instead, we study opaque location disclosures in a cyber-physical transport environment. Here location disclosure is not a means to an end, but rather part of a service that individuals cannot do without. To this end, our mixed-method research combines several existing scales on privacy concerns, cyber and physical risk taking, and transport choices (Section 3). We qualitatively analyse focus groups (Section 4) to explore the underlying motivations influencing location sharing and their impact on cyber-physical risk perceptions. We answer the following research questions:

RQ1: Do individuals cluster by their Internet privacy concerns and cyber and physical risk taking?

RQ2: How do the clusters compare in their experiences of location sharing?

RQ3: What can be learned to encourage adoption of future location based systems?

*Authors Becker and Posner contributed equally.

This research contributes a detailed understanding of how digital and physical concerns and behaviours coincide. The mixed-methods approach allows us to statistically examine behavioural constructs, and then explore how these manifest in location sharing risk perceptions and behaviours. This gives rise to several recommendations for policy (Section 5).

2. Related Literature

Many aspects of attitudes towards privacy have been studied. The privacy sub-type *Privacy of Location and Space* [6], which gives individuals the right to move in public without being tracked, identified or monitored is a particularly contentious issue nowadays. The authors consider various emerging technologies that impact on privacy; however, connected transport, self-driving cars and social fitness networks are not considered. The main research on location privacy is focused on either the advertising context [7], [8], check-in systems or Location Sharing Social Networks [9], [10], or anonymising location data through privacy enhancing technologies and statistical techniques [11].

However, we are interested in location privacy concerns in cyber-physical transport systems. This is an under-researched area, and no scales exist that can support participant segmentation. Moreover, transport is inherently physical, so it seems appropriate to view our research as a combination of Internet/information privacy concerns, and physical risk attitudes.

2.1. Privacy concerns

General privacy concerns is a well studied field. With existing scales, there is variation in the quality, application area and different approaches to measuring privacy concerns. Preibusch gives a comprehensive overview of the field [12], and recommends a number of validity-tested scales (such as [13]–[15]) that we will use in this research.

Earp *et al.* built and validated a construct on privacy values and concerns grounded in privacy policies [13]. The privacy protection goals underpinning the survey design are more akin to data protection principles, which fit well with our study's aim. The items were reviewed using survey experts and a panel of privacy experts, and pretested for reliability. Although designed to study e-commerce, the construct on Internet users' information privacy concerns (IUIPC) [15] is context-agnostic and widely used to research digital privacy concerns. In combination with the scale by Earp *et al.*, the questions give good coverage of data protection, privacy and Internet behaviour. Dinev *et al.* assemble another construct on privacy concerns in the Internet [14]. Their model incorporates the dimensions 'perceived vulnerability' and 'perceived ability to control information'. We found that the questions from Malhotra *et al.* [15] and Earp *et al.* [13] were more fitting for the context of our study, but that Dinev *et al.*'s questions [14] provided a useful basis for the development of our own items.

2.2. Existing privacy segmentation studies

There are a number of existing segmentation studies for privacy types. Historically, one of Westin's Privacy indices has been frequently used [16]. For example, Westin's Core Privacy Orientation Index segments participants into three groups: Privacy Fundamentalists, Privacy Unconcerned and Privacy Pragmatists. It has been used in mixed-methods studies on organisational, technological and environmental factors which influence technology adoption [17]. An alternative approach is described by Morton *et al.*, who use a Q methodology to segment participants based on the type of information cues they consider important [18]. The authors identify five groups: Information Controllers, Security Concerned, Benefit Seekers, Crowd Followers, and Organisational Assurance Seekers. Both Westin and Morton's approach to segmentation are valuable to their respective domains.

2.3. Location privacy research

The existing location privacy research primarily concerns itself with individuals' predispositions that influence their location privacy stance. Early studies concluded that people are not aware and do not care about being tracked [19], [20]. Krumm surveyed the literature in 2009 [20], and found that while research had identified the problem of location privacy and devised anonymisation techniques in response, people do not care about location privacy. Since then, researchers have identified that privacy expectations and concerns vary significantly by demographics and online activity [21], or by personal innovativeness [8]. Privacy concerns significantly influence continued adoption as compared to initial adoption [8]. There is no one fit-for-all solution for location based services.

2.4. Physical Risk

The Domain-Specific Risk Taking scale covers Ethics, Financial, Health/Safety, Recreational and Social aspects of risk [22]. Given the diverse physical risks of transport systems that can arise from location disclosure, this widely used and well established scale is highly appropriate for our research.

3. Survey study

Our aim in this part was to establish a robust grouping of participants which can be recovered through a small set of questions, rather than through an extensive survey. We conduct a comprehensive survey featuring a variety of relevant scales. We cluster the participants' responses, and identify a small set of survey questions that accurately predict the clusters. The survey study was conducted on LimeSurvey and participants were recruited using Prolific, a crowdsourcing platform. Based on the mid-2017 demographic distribution of the UK [23], we recruited 8 groups representative of the 18–24, 25–44, 45–64 and 65+ age brackets and two genders. All participants were required to be resident in the UK. This study was approved by the relevant ethics committee, and participants were shown an information sheet and gave informed consent for their participation. No personal identifiable information was collected as part of this study. The participants were reimbursed at a rate of £10 per hour and the survey took on average 17 minutes to complete.

3.1. Methodology

As discussed in the literature review, there are a large number of survey scales related to privacy. However, there is little in the domain of transport/location privacy. Instead of designing, testing and evaluating yet another scale, we chose to rely on a combination of existing, validated scales. An overview of the scales used can be found in Table 1. All Likert type questions were measured on a 7-point scale. The full survey text can be found in our supplementary material.

Table 1. Overview of survey questions.

Scale	# Q.	Source	Likert
Internet Users' Information Privacy Concerns (IUIPC)	12	[15]	agree
Information Sensitivity Scenarios	15	adapted from [15]	
Online Privacy Concerns	28	[13]	agree
Domain Specific Risk Taking (DOSPERT)	60	[22]	risky & likely
Online privacy relationship	18		
Transport choices	9		

The Internet Users' Information Privacy Concerns (IUIPC) construct [15] fits our research objectives well, as it focuses on technology related privacy concerns. It consists of three dimensions: collection, control and awareness, which are also applicable to the location privacy environment. Malhotra *et al.* also included two scenarios to measure the effect of the sensitivity of personal information requested [15]. Based on these, we created three scenarios for the location environment. While IUIPC focuses on the privacy concerns of users, the scale by Earp *et al.* includes potentially positive aspects of location data such as personalisation and participation of users [13].

In order to incorporate the physical aspects of location data, we included the Domain Specific Risk Taking (DOSPERT) scale [22]. Participants respond twice to 30 statements, once indicating how risky they perceive the statement to be, and once indicating how likely they are to engage in the described activity or behaviour.

We additionally introduced a new set of questions on privacy victimisation, and physical transport choices and concerns. These questions were developed in an iterative process with domain experts, and are similar to questions from [13], [24] and [14].

We carried out a power analysis for our segmentation analysis [25] before commencing data collection: taking IUIPC, Earp *et al.*'s scale, Dospert Likely, Dospert Risky, Transport Concerns, Transport Choices as dimensions, at least 420 participants are required to achieve 95% confidence in the segments [25]. We chose to recruit 520 participants. The survey included several attention checks. After removing participants who failed at least two, we ended up with 466 valid responses.

The majority of our participants walk or use cars as their main mode of transport, with 92% and 75% respectively using them at least once a week. Busses or trains are used much less frequently, with only 24% and 11% using them at least once a week. Ride sharing services were used infrequently, with 76% stating they never used them, and 13% stating they use them less than once a month.

3.2. Clusters of similar users

All responses to the scales outlined in Table 1 were clustered using k-means clustering after standardising and a dimensionality reduction using Principal Component Analysis (PCA). In PCA, the responses are transformed to a smaller space that preserves the variance while discarding noise. This improves the performance of the subsequent clustering, where the space is divided into k clusters of similar variance. Each transformed observation belongs to the cluster with the nearest mean. The number of clusters was chosen by evaluating the predictive capacity of a logistic regression classifier to recover the clusters, using 5-fold cross-validation. Four clusters had the highest accuracy at 92%. Visual inspection through silhouette analysis

Table 2. Descriptive statistics.

	CR	PR	TI	RA
Num. Participants	159	103	104	100
Average Age	53.5	51.0	42.4	36.5
Standard dev. Age	15.8	14.7	15.5	13.9
Percentage Male	34%	56%	44%	63%
IUIPC Collection	2.0	1.8^{*}	0.5^{**}	1.7
IUIPC Control	2.0	1.8	1.1^{**}	1.8^{*}
IUIPC Awareness	2.6	2.3	1.6^{**}	2.3^{**}
E. Personalization	1.9	1.6^{*}	0.0^{**}	0.9^{**}
E. Notice / Awareness	2.6	2.3^{**}	1.2^{**}	2.1^{**}
E. Transfer	2.7	2.7	1.3^{**}	2.3^{**}
E. Collection	2.4	2.2^{**}	0.7^{**}	1.4^{**}
E. Information Storage	2.5	2.2^{**}	1.4^{**}	1.9^{*}
E. Access / Participation	2.2	1.9^{*}	0.8^{**}	1.7
DOSPERT Likely	-1.6^{**}	-0.9^{**}	-1.3^{**}	-0.4
DOSPERT Risky	1.1	-0.4^{**}	0.2^{**}	0.0^{**}

Cyber risk takers (CR), physical risk takers (PR), transport innovators (TI), and risk abstainers (RA). Mean scores are presented for each scale. Within each row, **/* indicates a statistically significant difference at $p < 0.01/0.05$ between the distribution of responses of a given group compared to the distribution of responses with the group with the next largest mean (unpaired t-test). The four shades of cell background are used to denote the statistically significant ordering of the mean values in each row.

confirmed a clear separation of responses into 4 separate clusters. The responses from participants in these four clusters indicated the following four types: cyber risk takers, physical risk takers, transport innovators, and risk abstainers. Descriptive statistics for participants for each of these groups can be found in Table 2.

Our participants spread reasonably equally across the four types. There are minor demographic differences, with risk abstainers being younger and more male than other groups. The group identified as cyber risk takers is substantially more female.

Our clusters split the existing scales into statistically significant segments, highlighting the link between their discriminant validity and ours. The transport innovators score statistically significantly lowest on all IUIPC and Earp *et al.* factors [13], [15]. However when considering physical risks, the transport innovators are more middling. Conversely, the cyber risk takers consistently score highest across all IUIPC and Earp *et al.* factors. These participants are the most concerned about all aspects of information privacy, and want full control over their data. At the same time they are averse to physical risk. Physical risk takers are, as expected, least likely to consider the DOSPERT statements to be risky, while being most likely to undertake them.

3.3. Reduced survey set

It is infeasible to use a 142 question survey to cluster users in subsequent studies or real-life applications. Therefore, we reduced the number of questions required to group the participants into the 4 clusters from 142 to 17. The literature describes a variety of methods. Essentially each of our questions is treated as a feature, and the task becomes one of feature selection. Several methods were attempted, such as correlation based feature selection [26], feature importance ranking with random forest trees, and mutual information, chi-squared and ANOVA F-test based feature selection. We compared the performance of these in a 5-fold cross validated setting of the predictive capacity of a logistic regression classifier to recover the clusters for features in the range between 1 and 142. Recursive feature elimination performed best and achieved an accuracy score of 0.75 with 17 questions which was deemed an acceptable trade-off between numbers of questions and accuracy by the authors.

The full reduced question set can be found in the supplementary files. Two of the questions are drawn from IUIPC [15], five from Earp *et al.* [13], and two stem from our own questions. Two and five statements remain for the DOSPERT Likely and Risky scales respectively [22]. The final question is scenario three, where participants respond on a 7-point Not probable/probable scale.

3.4. Discussion

Answering RQ1, this research has demonstrated that there are distinct privacy types in the location privacy environment. These types are not demographic artefacts, but rather describe inherently different attitudes. The types partially align with the existing privacy constructs, but are extended through a combination of physical risk perceptions and risk taking. Each cluster has distinct properties that make them suitable for tailoring interventions and messaging specifically at them. The reduced question set makes it possible to accurately classify an individual into one of these four clusters with only 17 questions.

4. Focus groups

This section qualitatively explores the underlying motivations influencing location sharing and their impact on cyber-physical risk perceptions, and investigates acceptable options to reducing these risks.

4.1. Methodology

Participants were recruited through our organisation's participant pool. This database includes details of over 2000 residents in the area who had registered an interest and consented to be contacted for research purposes. The reduced survey set described previously was used for pre-screening. Each participant was assigned to one of the four privacy types. For each of the 4 types, we conducted 3 separate focus groups with 6 participants per group. The study was reviewed and received ethics approval.

4.1.1. Design Focus groups offer a comfortable environment to promote discussion and allow participants to discuss their views, thoughts and beliefs on the topic. We used homogeneous groups (same attitude to risk cluster) of strangers, which are favoured when exploring personal and sensitive topics such as privacy as those taking part have similarities as well as having their own experiences and opinions regarding the topic discussed.

The topic guide was split into four key sections. The first focused on the motivations that influenced location sharing on transport-based apps, the second on the perceived benefits and risks associated with location sharing, the third on the possible relationship between location sharing and cyber-physical crime, and the fourth on identifying the tools that would support members of the general public to make safer and more informed decisions when choosing to share location. We did not collect any other demographics, or information on our participants' use of transport apps or transport choices.

4.1.2. Procedure After being identified through their survey responses, participants were sent an email inviting them to take part in a focus group. The email included the consent form and participant information sheet. The focus groups took place at our organisation, and lasted approximately 90 minutes. Focus groups were recorded for transcription. At the beginning of each focus group, participants were once again presented with the information sheet and consent form and reminded of their right to withdraw. Participants were reminded of the nature of the study, its confidentiality and anonymity. At this point recording began. Two researchers were present during each focus group. One facilitated the focus group while the other took notes. These were also used in the analysis of the data. Once the focus groups were completed participants were provided with a full debrief and the opportunity to raise any further questions. Participants were given a £25 incentive.

4.2. Data analysis

Upon completion the focus groups were transcribed and analysed by one researcher using inductive thematic analysis via NVivo. Thematic analysis provides a strong analytical tool to identify, analyse and report patterns within a data set, a key criterion due to the exploratory nature of this study. In addition, its independence from any theoretical framework makes it a strong analytical tool for this study. An inductive approach was chosen to ensure that the findings were based entirely upon the data and generated by the respondents themselves, not driven by any previous theoretical ideas. The chosen methodology allowed for an in-depth and detailed analysis of the participants' own thoughts within this novel area. After coding two out of the three focus groups for each cluster, saturation was reached; the third focus groups added only minor nuances.

4.3. Results: common themes

The themes discovered in the focus group and their occurrence between the groups can be seen in Table 3. In response to RQ2, many of the themes support our clustering and the participants' responses to the segmentation scales: for example, both types of risk takers believed that location disclosures improve the safety of the system to users. All clusters believed that age would influence location sharing decisions, but interestingly, the age distribution between the clusters is not statistically significantly different. Transport innovators were positive about the consequences of location sharing, believing that it offered improved choice, better accountability, and more accurate information provided by systems. The following sections will discuss the commonalities and differences of these groups that surfaced during the focus groups.

4.3.1. Age All clusters believed that age would influence location sharing decisions with different generations choosing different settings, predominantly because they had different perceptions of risks:

> *'That's an age thing, isn't it, as well, for us? These guys, it's automatic. For us, we make it more complicated than it probably is.'*

With younger generations growing up immersed in technology it was believed that they knew more about data sharing and therefore would be less susceptible than older generations to fall into some of the traps set out by perpetrators. However, while younger generations may be more immersed in technology and have an

Table 3. Focus group responses for different clusters.

Clusters				Factors influencing location sharing decision
CR	PR	TI	RA	
✓			✓	Safety (improving)
	✓	✓		Safety (reducing): includes physical safety (attacks, burglaries)
✓			✓	Value attributed to data (lack of)
		✓		Security of the device itself
			✓	Defeatist towards preventing cybercrime
✓	✓	✓	✓	Age: generational differences in how location was shared and why
		✓		Job
✓				Ease of use
✓	✓	✓		Convenience and frustration with app (hassle of entering location manually)
✓		✓		Functionality
		✓		Accuracy and real-time information provision
✓	✓	✓	✓	Trust in the app provider (or lack of)
✓		✓	✓	Transparency of the app provider (what is collected, why, how and for whom)
✓		✓	✓	Privacy: particularly loss of privacy
	✓	✓	✓	Providing and receiving a service/benefit
✓			✓	Uncertainty of future transport systems
	✓			Mood
	✓			Environment
	✓		✓	Storing vs data sharing
			✓	Ability to stay anonymous when sharing
	✓	✓	✓	Having a choice in the decision of whether their data is shared or not
	✓	✓		Size of audience (lack of understanding)
	✓	✓	✓	Apps' reason for accessing location
		✓	✓	Personal bad prior experience
✓	✓	✓	✓	Recommendations and experiences from family and friends
✓				Recommendations from online networks
✓			✓	Reputation of app provider (history of data misuse or breaches)
		✓		Accountability of the app provider in the event of something going wrong
✓	✓	✓	✓	Lack of understanding of the potential risks associated with location sharing and new forms of crime
			✓	Vulnerability of the network through which data is shared

cyber risk takers (CR), physical risk takers (PR), transport innovators (TI), and risk abstainers (RA).

overall awareness of the risks of location sharing there was an overall belief that younger generations were too naïve when it came to data sharing as a result of their familiarity with these technologies and consequently did not appreciate the risks.

The perceptions on age shown in the focus groups appear contradictory to our clusters, where cyber risk takers had the highest mean age, and the risk averse cluster had the lowest mean age. However this may be false confidence: rather than being risk averse, the older generation appears to be ignorant of the risks. The cyber risk takers judge physical risk highest and are least likely to do risky activities. However, this behaviour does not translate into online risk taking.

4.3.2. Trust (or lack of) in the app provider Trust was another important factor across all of the clusters, and particularly trust in the app provider. Trust was quite a complex notion and was made up of several different elements which included whether the app provider was a recognised name. All the clusters, apart from the physical risk takers, felt more comfortable sharing location data with a recognised 'household name' than a new start-up.

> *'We trust British Airways, for whatever reasons, it's institutional names, recognised and it's safe. I would with them, that wouldn't worry me, it's a trusted organisation.'*

However, physical risk takers were less likely to trust big organisations and therefore share their location with them. This was due to the frequent media reports surrounding the misuse of data by bigger organisations leading to them questioning what was 'happening in the background'.

Friends and family, who they believed to have their best interest at heart, were seen as the only reliable source. Only the cyber risk takers also used reviews from online communities even though they were aware that these 'had to be taken with a pinch of salt' and should be read as part of a wider context:

> *'Yes, friends and family, because you would hope that they're not going to recommend something or someone dodgy, so yes, you would always, I think, go with their recommendation.'*

4.3.3. Transparency of the app provider Across all groups this was one of the most important factors that influenced their decision making process when choosing to share location. Organisations that were open about why they were accessing the information, why they requested location sharing, how the data was being used, who was accessing the data, and how the data was shared were ones that users felt more comfortable sharing location with.

When faced with the prospect of ever more connected transport systems, all users demanded more

transparency, with some going as far as saying that this should be a legal requirement. Indeed, participants believed that increased transparency from the app provider about why they are collecting the data, what data they are collecting, and for whom would encourage more data sharing on the user's part:

*'So, it's a question of information, I think, going forward, without pages and pages and pages of s**** to read and you go, 'yeah', but easy to understand. Where is data stored, what do they do with it, what criminals are out there, how can they use it, that sort of thing.'*

4.3.4. Safety All clusters mentioned safety as an important factor when choosing to share location information, but there was a divide across the clusters in terms of those who believed that location sharing could improve safety and those who thought that overall it would lead to reduced safety. While all clusters recognised the possible risks associated with location sharing, cyber risk takers and risk abstainers saw sharing location as having a number of safety benefits, particularly towards ensuring the safety of other. Members of these clusters consistently reported that location sharing provided 'reassurance', allowing them to know the whereabouts of children, particularly children with special needs, family members and generally more vulnerable members of society (older adults, especially ones with more neurological disabilities).

'I'd say it'd be useful, I mean, I do some voluntary work with dementia groups, and some of those poor people, I think, that app, to track them, if they went missing from home and you didn't know where they were, that would be an absolutely fantastic app.'

These clusters were very realistic about the potential risks associated with location sharing, however there was an overall agreement that there was already so many other things to worry about without having to worry about location sharing. Therefore when choosing to share location users worked out the degree of risk associated with sharing their data, whereby they weighed their perceived risks and benefits of sharing location and made their decision from there. If the risks, which tended to be linked to physical crimes (such as being attacked or burglaries), cyber-crimes (identity theft, financial crime) and cyber-physical crime (cyber and physical stalking) outweighed the benefits significantly then they may reconsider sharing their location.

'I work out what risk am I actually running by giving someone my location, and I think it's so minimal, I don't worry and I share it.'

4.4. Results: differences between clusters

Here we discuss some of the differences between clusters in more detail.

4.4.1. Future transport systems Cyber risk takers discussed several concerns over future types of crimes and particularly in relation to connected vehicles and future transport systems more generally. One of the main concerns was the ethics surrounding these types of vehicles, particularly how they would act in the event of an road collision, leading to a lack of trust:

'Who's culpable in a driverless car, if something happens?'

Members of this cluster reported concerns over what data would be required, who would be able to access this data, how the data would be used and how the transport systems would work. This lack of certainty often lead to concerns over location sharing in these new forms of transport as it was unclear what the possible risks could be. However, there was also an awareness of the many benefits that could come with increased location sharing in future transport systems. The increased connectedness of the transport systems might lead to them being easier to navigate, making travelling faster and more accessible, as well as safer. However, it was evident that for this group the risks of location sharing on future transport systems was not yet understood:

'Human error has to account for an awful lot of accidents, so if you took away human error and human stupidity, probably in the long run, the automated vehicle would probably cause a lot less injuries than humans do, to be fair, but it's still that, you just don't know enough about all of it.'

4.4.2. Storing vs sharing data For physical risk takers there seemed to be a clear distinction between sharing and storing data. While this group may feel comfortable sharing their location in certain instances, they did not want this data to be stored. The storing element seemed to be the one that was associated with the most risks with some users saving incorrect addresses (for example postcode only or former addresses) in their vehicles or apps in order to avoid their personal information being misused in the even of their vehicles being stolen or their phone being hacked.

'My Sat Nav in my car, I think if my car's stolen, I don't want people knowing my actual address, so I just put it into the postcode, so it could be anywhere. I don't mind sharing information but I don't want them to have it, if that makes sense'

4.4.3. Accuracy and real-time information provision While many other clusters shared location with the aim of receiving a service, transport innovators paid particular attention to the information provision that came with location sharing. They perceived the increased accuracy of information due to location sharing as one of the highest benefits, particularly when using a navigation tool.

'Some of the apps, if you're looking for somewhere in the locality you are, yes, it's great, just share your location, if you're out and about, it's fine, and it'll come up with, I don't know, any sort of shops or whatever you want, because that's very convenient'

They were aware that sharing location was associated with potential risks, but these risks were considered as worthwhile if it meant that real-time information provision could be obtained.

4.4.4. Accountability of the app provider One of the concerns of transport innovators was the lack of accountability if there was a data breach, or data was misused.

'I'm not saying it's easy but I think, as I said, we need to start holding people accountable for some of the things that they're doing.'

They reported that this was a barrier to sharing location with certain apps, particularly ones that may have previous incidences of data breaches and failing to take accountability for these. Members of this cluster reported that they would be more willing to share location with an app if there was someone that they could seek information from and that would be held accountable in the event that something went wrong. They reported that this accountability should be something required from all organisations and a priority for future transport systems.

'Something has to be done to hold them to account going forward. That should be a priority.'

4.4.5. Attitude towards preventing cybercrime Unlike other clusters that were aware of the potential risks associated with sharing location, risk abstainers had a very defeatist attitude towards being able to prevent perpetrators from committing crimes:

'I'm sure there are people out there that if they want it, it's like breaking into cars, you can have the security in the world, but if they want something in there, they'll get in there, like your house or your data.'

Indeed, this group were unsure whether as a society we were able to cope with these new forms of crimes and consequently the new types of perpetrators that were emerging (e.g. hackers). They reported that perpetrators were smart and continuously getting more sophisticated, making it hard to see what could actually be done to avoid being a victim of these new types of crime and prevent them more generally. There was an overall understanding that perpetrators are always getting smarter and therefore it was always about 'staying one step ahead for as long as possible'.

'But all the hackers are always one step ahead of the Facebook and the Instagram.'

This group believed that their data was not particularly valuable, therefore they had no problem with sharing their location. Further they felt that if an organisation or a perpetrator wanted to access their data, they could do nothing about it.

4.4.6. Ability to stay anonymous While other clusters discussed the differences between sharing and storing data, risk abstainers placed particular importance on the ability to stay anonymous when sharing their data. They reported that if they were able to share their data anonymously, with none of their data actually being traceable to them they would be more willing to share their location. This was especially the case if it could help improve services overall.

'If your identity and your very sensitive material remains anonymous, then I'm happy for me to be, let's say, more of a statistic, especially if that then goes on to provide more useful information to help us get round. Great, but yes, I would much rather that they don't know who it is exactly.'

5. Discussion

This part of the study aimed to explore in more detail the factors that influenced location sharing decisions across four distinct clusters, particularly the similarities and differences across each of these groups.

Overall the qualitative analysis highlighted a number of important differences across each of the clusters, providing further support for the four distinct groups that were identified. These differences demonstrated varying approaches to location sharing. Some clusters

placed greater weight on the possible benefits, which ranged from providing a service to a community to improving the accuracy of the information that they could obtain. Other benefits were very much grounded in improved safety, whereby location sharing could ensure the security and safety of some of the more vulnerable members of society. On the other hand, some were very concerned about sharing their location as they believed that despite the possible benefits, they were far outweighed by the risks and the new types of crime that could emerge as a result of increased location sharing (for example car-hacking, burglaries, and crimes that happened across both the cyber and physical domains such as stalking/cyber-stalking).

In response to RQ3, these differences in the values influencing location sharing provide further support for the need for a more tailored and personalised approach to information provision. Failing to recognise these differences in motivations could lead to reduced data sharing or more risky sharing behaviour by some clusters.

While there were differences across these clusters, there were also a number of motivations that were common across all four. Participants agreed on the importance of transparency from those requesting their data. The lack of understanding over how data was currently being used, why it was being collected, who could access it and how it was being shared was a particular concern. Overall, the clusters all acknowledged that location sharing would be a necessity for future transport systems, however many felt resistant to making this change, in part due to the feeling of 'having no say in the matter', and app providers or data controllers failing to provide the information they required in an accessible way. Our participants essentially resign themselves to the privacy paradox as they accept the seemingly necessity to surrender their location privacy [27]. At no point did the discussion touch upon GDPR or other existing legislation.

Behaviour change is an extremely complex process and as research has consistently demonstrated, attitudes are not the strongest predictor of future behaviour, but are only one of the factors influencing behaviour. Most models and the behavioural literature more widely emphasise that relying on attitudes to explain behaviour has several limitations. For example, cognitive dissonance theory [28] highlights the discrepancies between attitudes and behaviour, including in relation to privacy behaviour [29]. The qualitative work carried out as part of this research has allowed us to address these limitations by exploring in much more depth the values and motivations influencing behaviour and location sharing behaviour more specifically.

5.1. Implications for policy

The findings from this research have an number of impacts on potential future policies. The evidence provides support for a more tailored and personalised approach to information provision surrounding location sharing in future transport systems, including transport apps and websites. Failing to recognise these differences could lead to reduced data sharing, riskier sharing behaviour increasing the likelihood of users being victims or perpetrating cyber-physical crimes or even total avoidance of new forms of transport.

One of the key implications for policy is the need for increased transparency from app providers as well as transport providers if they want to ensure the retention and safety of their users. In order to achieve this, and help rebuild trust in app providers, many participants suggested the need for better standards and independent regulatory bodies to uphold those standards. In the event of a crime (physical or cyber) that was promoted by these tools, those responsible need to be held to account. In addition, continuous education tailored to the needs of the different segments should be considered. Finally, further research should be conducted to identify the benefits of tailored information provision in supporting informed location sharing decisions, and the most successful medium through which this can be achieved.

6. Conclusion

This research contributes a novel understanding of privacy concerns in the area of cyber-physical systems. Based on existing scales both from cyber and physical areas and a UK representative study with 466 participants, we have identified a new set of location privacy types that can classify an individual accurately in just 17 questions. Through a mixed-methods study design, we explored these types in more detail through homogeneous focus groups, identifying differences in risk appetite and risk understanding. Age of individuals, trust and transparency (or lack thereof) of providers as well as safety implications are common factors perceived to influence location sharing decisions of all clusters. Cyber risk takers are distinctive, as they prefer convenience and do not value their data highly. Despite this, they are apprehensive of physical risks. Physical risk takers are more influenced by their mood and environmental factors, and struggle with the perpetuity of shared data. Transport innovators stand out through their dedication to technology. They experiment, and believe in the accountability of providers. Risk abstainers had the youngest average age in our sample. They are defeatists, and will only share their data if

they remain anonymous and can trust the security of the infrastructure.

These differences of opinion will give rise to diverse demands on providers and legislators. Public services need to be accessible to everyone, so it is essential that expectations are incorporated into design processes and that individuals' rights are respected. If service providers wish to attract users from these four diverse clusters they will need to address each groups' needs and preferences in a targeted manner.

Data Availability

The full and reduced survey set and the analysis files can be found at `https://github.com/watercrossing/privacy-in-transport`.

Acknowledgements

This work was supported by the EPSRC project "ACCEPT: Addressing Cybersecurity and Cybercrime via a co-Evolutionary aPproach to reducing human-relaTed risks", grant number EP/P011896.

References

[1] N. Rose. (2018). Strava released their global heatmap, [Online]. Available: `https://twitter.com/nrg8000/status/957318498102865920`.

[2] E. Uhlemann, *'Connected-vehicles applications are emerging [connected vehicles]'*, *IEEE Vehicular Technology Magazine*, vol. 11, no. 1, 2016.

[3] Office for National Statistics, Overview of burglary and other household theft: England and Wales, 2017.

[4] Burgess, Kaya, *'Thieves 'followed rider on his strava app' to make off with £12,500 in bikes'*, *The Times*, 2018.

[5] S. Parkin, T. Patel, I. Lopez-Neira and L. Tanczer, *'Usability Analysis of Shared Device Ecosystem Security: Informing Support for Survivors of IoT-Facilitated Tech-Abuse'*, in *NSPW*, 2019.

[6] R. L. Finn, D. Wright and M. Friedewald, *'Seven Types of Privacy'*, in *European Data Protection: Coming of Age*, S. Gutwirth, R. Leenes, P. de Hert and Y. Poullet, Eds., Dordrecht: Springer Netherlands, 2013.

[7] S. Banerjee, *'Geosurveillance, Location Privacy, and Personalization'*, *Public Policy & Marketing*, vol. 38, no. 4, 2019.

[8] H. Xu and S. Gupta, *'The effects of privacy concerns and personal innovativeness on potential and experienced customers' adoption of location-based services'*, *Electron Markets*, vol. 19, no. 2, 2009.

[9] J. Lindqvist, J. Cranshaw, J. Wiese, J. Hong and J. Zimmerman, *'I'm the Mayor of My House'*, in *SIGCHI Conference on Human Factors in Computing Systems*, Vancouver, BC, Canada: ACM, 2011.

[10] X. Page, B. P. Knijnenburg and A. Kobsa, *'FYI: Communication Style Preferences Underlie Differences in Location-sharing Adoption and Usage'*, in *2013 Conference on Pervasive and Ubiquitous Computing*, Zurich, Switzerland: ACM, 2013.

[11] K. P. N. Puttaswamy, S. Wang, T. Steinbauer, D. Agrawal, A. E. Abbadi, C. Kruegel and B. Y. Zhao, *'Preserving Location Privacy in Geosocial Applications'*, *IEEE Transactions on Mobile Computing*, vol. 13, no. 1, 2014.

[12] S. Preibusch, *'Guide to measuring privacy concern: Review of survey and observational instruments'*, *International Journal of Human-Computer Studies*, vol. 71, no. 12, 2013.

[13] J. Earp, A. Anton, L. Aiman-Smith and W. Stufflebeam, *'Examining Internet Privacy Policies Within the Context of User Privacy Values'*, *IEEE Transactions on Engineering Management*, vol. 52, no. 2, 2005.

[14] T. Dinev and P. Hart, *'Internet privacy concerns and their antecedents - measurement validity and a regression model'*, *Behaviour & Information Technology*, vol. 23, no. 6, 2004.

[15] N. K. Malhotra, S. S. Kim and J. Agarwal, *'Internet Users' Information Privacy Concerns (IUIPC): The Construct, the Scale, and a Causal Model'*, *Information Systems Research*, vol. 15, no. 4, 2004.

[16] P. Kumaraguru and L. Cranor, *'Privacy indexes: A survey of Westin's studies'*, Carnegie-Mellon University, Technical Report 856, 2005.

[17] A. Morton, *'"All my mates have got it, so it must be okay": Constructing a Richer Understanding of Privacy Concerns—An Exploratory Focus Group Study'*, in *Reloading Data Protection*, Springer, Dordrecht, 2014.

[18] A. Morton and M. A. Sasse, *'Desperately seeking assurances: Segmenting users by their information-seeking preferences'*, in *2014 Twelfth Annual International Conference on Privacy, Security and Trust*, 2014.

[19] E. Kaasinen, *'User needs for location-aware mobile services'*, *Pers Ubiquit Comput*, vol. 7, no. 1, 2003.

[20] J. Krumm, *'A Survey of Computational Location Privacy'*, *Personal Ubiquitous Comput.*, vol. 13, no. 6, 2009.

[21] Z. Gardner, D. Leibovici, A. Basiri and G. Foody, *'Trading-off location accuracy and service quality: Privacy concerns and user profiles'*, in *International Conference on Localization and GNSS*, 2017.

[22] A.-R. Blais and E. U. Weber, *'A domain-specific risk-taking (DOSPERT) scale for adult populations'*, *Judgment and Decision Making*, vol. 1, no. 1, 2006.

[23] Office for National Statistics, Estimates of the population for the UK, England and Wales, Scotland and Northern Ireland, Mid-2017, 2018.

[24] H. J. Smith, S. J. Milberg and S. J. Burke, *'Information Privacy: Measuring Individuals' Concerns about Organizational Practices'*, *MIS Quarterly*, vol. 20, no. 2, 1996.

[25] S. Dolnicar, B. Grün, F. Leisch and K. Schmidt, *'Required Sample Sizes for Data-Driven Market Segmentation Analyses in Tourism'*, *Journal of Travel Research*, vol. 53, no. 3, 2014.

[26] S. Chormunge and S. Jena, *'Correlation based feature selection with clustering for high dimensional data'*, *Journal of Electrical Systems and Information Technology*, vol. 5, no. 3, 2018.

[27] S. Kokolakis, *'Privacy attitudes and privacy behaviour: A review of current research on the privacy paradox phenomenon'*, *Computers & Security*, vol. 64, 2017.

[28] L. Festinger, A Theory of Cognitive Dissonance. Stanford university press, 1962, vol. 2.

[29] S. B. Barnes, *'A privacy paradox: Social networking in the United States'*, *First Monday*, vol. 11, no. 9, 2006.