"I didn't know I was that much of a bad boy": Surprise about Driving Data

Maria Hyun
University of Wisconsin-Madison
mhyun6@wisc.edu

Eve He
University of Wisconsin-Madison
eve.he@wisc.edu

Emilee Rader
University of Wisconsin-Madison
ejrader2@wisc.edu

1 Introduction

Driving a car is an important part of daily life for most people in the United States [14]. However, people have a limited understanding of the privacy risks associated with data collected in an automotive context [2, 3, 8]. Cars have more different kinds of sensors than smartphones do, because they collect and store detailed performance data for use during diagnostics or after an accident. Cars also collect very sensitive information, about how people break the law (e.g., speeding), where they go and when, and cars even collect data on how much passengers weigh [5]. Auto manufacturers and dealerships do a poor job of providing information at the time a car is purchased about what kinds of data will be collected by the car, who will have access to that data, and how it will be used, and owners manuals also rarely include this information [9].

Recent research has emphasized the role of surprise about data collection and inferences in privacy-related contexts [1, 6, 12, 13]. Surprise occurs when a person's expectations are violated [11], and signals unanticipated data practices. The goal of this research was to investigate people's awareness and perceptions of data collected about them by their cars, to better understand how their existing knowledge might affect how they think about driving data privacy. This poster presents preliminary findings from interviews focusing on participants' reactions to data collected about their driving over a 12 week period, highlighting aspects of the data they found surprising. People cannot make informed privacy decisions about data and inferences they are not aware of, and so identifying when and why surprise occurs can help privacy designers create interventions targeted towards gaps in user knowledge.

USENIX Symposium on Usable Privacy and Security (SOUPS) 2025.
August 10–12, 2025, Seattle, WA, United States.

2 Method

We collected driving data from 40 participants during October 2020-March 2021. Participants were recruited via a university-run subject pool which solicited members from the broader community. While these data are 4 years old, people in the U.S. keep their cars for 12.5 years on average [10], indicating that adoption of new car-based technologies happens more slowly than for other personal computing technologies.

The driving data were collected by a device plugged into each car's diagnostics port. As part of the consent process, participants were informed that the system would collect data such as car speed, braking, acceleration, trip times, and geolocation. After 12 weeks, a report was generated for each participant consisting of visualizations of their driving data, along with aggregated trends across the entire dataset and data from third party APIs (e.g., nearby crime activity, Google Street View images, deviation from posted speed limits). More details about the method are available in Hautea et al. 2021 [7].

This poster focuses on preliminary qualitative analysis of 39 interviews[1] that were conducted as participants reviewed their driving data report for the first time. The interviewer asked follow-up questions probing for participants' understanding of and reactions to the data. Interviews were recorded and manually transcribed, and analyzed using an iterative inductive qualitative process in which two members of the research team coded for emerging themes. This study was approved by the relevant institutions' IRBs.

3 Findings

What was not surprising? The element of surprise is a crucial part of our research, as it prompts participants to reconsider data collection and driving behavior. However, not all data were surprising; some simply confirmed participants' expectations. Many anticipated that the data would include distance, speed, and braking. P05, for instance, expected this type of data and was not surprised, as she was already familiar with driving data through a friend's insurance program.

[1] One participant dropped out of the study before being interviewed.

> "I assumed it would break down all my trips, track all my trips that I took, and I thought there would be some things about braking and acceleration because my friend said that's part of what her insurance program tracks." *–P05, Female, 54 y.o.*

Matching with their own recollection is another reason why participants were not surprised by their data. P09 was not surprised, as she already knew Christmas is a busy time for family visits and expected a high amount of driving.

> "Over Christmas week, I took 17 trips of 127 miles. [...] I'm not surprised that I still did a bunch of driving during Christmas." *–P09, Female, 29 y.o.*

Surprise due to mismatch with memory. Participants were surprised when what they remembered did not match what was in their report. This mismatch was most common in three areas: hard braking, speeding, and driving frequency. For some, the total number of trips felt unexpectedly high, even though they knew they drove regularly.

> "Oh, 288 trips in my car. That feels like a lot in 12 weeks. I mean, I know I drive a fair amount, and then getting kids to school and from school, and driving to pick up random things and stuff like that. But that still feels like a lot." *–P41, Female, 41 y.o.*

For these people, the surprise came from seeing their driving data counted up in a way that made them seem more frequent than they had remembered.

Other participants expressed surprise at the speeding data in their reports, including how high their top speeds were and how frequently they had exceeded the posted limits. P03, who admitted to sometimes speeding, was still very shocked to see several data points showing speeds more than 30 miles per hour over the limit. Some participants initially doubted the speeding data, but after reflecting they recalled specific incidents, which led to recognition and eventually self-correction.

> "I didn't know I was that much of a bad boy... Speeding over the limit because I never thought I was speeding over the limit." – *P19, Male, 80 y.o.*

Similarly, some were surprised by how many hard brake events were recorded. While these participants knew they sometimes did this, they did not agree with the high number of hard brakes in the report. P35 initially expressed disbelief, but then considered what it might mean about his driving.

> "56 hard brakes. What? That's crazy. There's so many. I didn't think it would be that high. [...] The hard braking really gets me, because I always thought I was a really good driver, and I still do... Hard braking is bad, I mean, it's bad for the people behind me."*–P35, Male, 35 y.o.*

Surprise due to exposing lifestyle patterns. Participants were also surprised by how data about driving reflects broader life patterns such as routines, habits, and frequently visited locations. P08, who had focused only on driving data, was surprised the device could access and present so much personal information.

> "It amazes me that all this can be collected [...] you can really pinpoint a lot of information about a person's habits in a week's time, or even in a day if they're a daytime traveler, nighttime traveler. You can get a good idea of how they work nights or what their lifestyle is like."*–P20, Female, 49*

Finally, one thing that surprised participants was the level of detail in location data associated with their driving activity. P30 expressed that she did not realize the implications of collecting location data as part of the research.

> "I didn't know it was going to be really keeping track of all the locations and the number of times, visiting each location and the time of day and all that. The amount and the time it took, et cetera." *–P30, Female, 55 y.o.*

4 Discussion

Our findings show that participants were surprised by their driving data when it did not match what they remembered about their driving, and when it revealed broader patterns about their lives. Memory is a representation of the past, as is the computational output generated by sensors and algorithms. Each is biased, and our findings suggest that they may be biased in different ways.

For example, driving requires a high level of concentration, and people's memory for everyday driving events is poor [4]. It is trivial for a machine to aggregate data over time to reveal patterns, but hard for people to do so. This may make it more difficult for people to notice or remember their actions in the same way as an automated data collection system can. People may recall speeding as a general practice, whereas speeding data can pinpoint in specific detail moments where and when speeding was taking place and by how many miles per hour. Similarly, hard braking as measured by exceeding an accelerometer threshold does not account for situational factors that may be more memorable, like how an accident was avoided. As a result, the driver's memory of the event may not even include their perception of the acceleration forces.

Surprise signals an expectation violation due to unanticipated data practices. If aspects of the task or context make it harder for a person to notice or perceive events in the same way the technology can, surprise may occur. This is problem for privacy, because people cannot make up-front privacy decisions about data and inferences they do not expect. Transparency interventions to help people make better privacy decisions about their driving data should take into account inherent differences between how people and machines keep track of and aggregate information. They should also adapt to situational factors, such as the effort and focus required to drive a car safely, that could make it more difficult for the human and system perspectives to agree.

Acknowledgments

We thank Joe Freedman, Norbert Nthala and Rick Wash for their assistance with developing the software for this project. We also thank Samantha Hautea and Faye Kollig, and the BITLab @ MSU research group for their contributions and feedback. This material is based upon work supported by the National Science Foundation under Grant No. CNS-1524296.

References

[1] Patricia Arias-Cabarcos, Saina Khalili, and Thorsten Strufe. Surprised, shocked, worried: User reactions to Facebook data collection from third parties. *Proceedings on Privacy Enhancing Technologies (PoPETs)*, 2023(2):215–234, 2023.

[2] Giampaolo Bella and Pietro Biondi. Car drivers' privacy awareness and concerns, 2023. https://www.researchgate.net/publication/374001207_Car_Drivers'_Privacy_Awareness_and_Concerns.

[3] Giampaolo Bella, Pietro Biondi, and Giuseppe Tudisco. Car drivers' privacy concerns and trust perceptions. In *Trust, Privacy and Security in Digital Business: 18th International Conference, TrustBus 2021, Virtual Event, September 27–30, 2021, Proceedings 18*, pages 143–154. Springer, 2021.

[4] Samuel G. Charlton and Nicola J. Starkey. Memory for everyday driving. *Transportation Research Part F: Traffic Psychology and Behaviour*, 57:129–138, 2018. Special Issue on Everyday Driving.

[5] Bill Hanvey. Your car knows when you gain weight. *New York Times*, May 20, 2019. https://www.nytimes.com/2024/03/11/technology/carmakers-driver-tracking-insurance.html.

[6] Samantha Hautea, Anjali Munasinghe, and Emilee Rader. 'That's not me': Surprising algorithmic inferences. In *Extended Abstracts of the 2020 CHI Conference on Human Factors in Computing Systems*, CHI EA '20, page 1–7, New York, NY, USA, 2020. Association for Computing Machinery.

[7] Samantha Hautea, Norbert Nthala, Faye Kollig, Joao Marcelo Ferraz, and Emilee Rader. "Assertive driver, I can imagine that": Interpretations of inferences from driving data. In *Poster at the Symposium on Usable Privacy and Security*. USENIX Association, 2021.

[8] Kashmir Hill. 'Smartphones on Wheels' draw attention from regulators. *New York Times*, April 30, 2024. https://www.nytimes.com/2024/04/30/technology/regulators-investigate-carmakers-driver-tracking.html.

[9] Kashmir Hill. Automakers are sharing consumers' driving behavior with insurance companies. *New York Times*, March 11, 2024. https://www.nytimes.com/2024/03/11/technology/carmakers-driver-tracking-insurance.html.

[10] Paul Lienert. U.S. consumers keep vehicles for a record 12.5 years on average. *Reuters*, May 15, 2023. https://www.reuters.com/business/autos-transportation/us-consumers-keep-vehicles-record-125-years-average-sp-2023-05-15/.

[11] Emilee Rader. Examining user surprise as a symptom of algorithmic filtering. In *International Journal of Human Computer Studies*, volume 98, pages 72 – 88, 2017.

[12] Emilee Rader, Samantha Hautea, and Anjali Munasinghe. "I have a narrow thought process"': Constraints on explanations connecting inferences and Self-Perceptions. In *Sixteenth Symposium on Usable Privacy and Security (SOUPS 2020)*, pages 457–488, 2020.

[13] Emilee Rader and Janine Slaker. The Importance of Visibility for Folk Theories of Sensor Data. In *Symposium on Usable Privacy and Security*, 2017.

[14] Katherine Schaeffer. 1 in 10 americans rarely or never drive a car. *Pew Research Center*, November 14, 2024. https://www.pewresearch.org/short-reads/2024/11/14/1-in-10-americans-rarely-or-never-drive-a-car/.