"Assertive driver, I can imagine that": Interpretations of Inferences from Driving Data

Samantha Hautea
Michigan State University
hauteasa@msu.edu

Norbert Nthala
Michigan State University
nthalano@msu.edu

Faye Kollig
Michigan State University
kolligfa@msu.edu

João Marcelo Ferraz
Universidade Federal de Pernambuco
joao.pferraz@ufpe.br

Emilee Rader
Michigan State University
emilee@msu.edu

Abstract

Privacy in sensor-based systems must be studied in the context of lived experiences to understand how behavior may differ from stated preferences. In this study, we showed participants inferences about their driving style that we generated from data about their actual driving behavior, and qualitatively analyzed their reactions. Preliminary findings suggest most participants rationalized the inferences by making them fit within their positive image of their own driving. This suggests that people may find it difficult to conceptualize uses of inferences that run counter to their self-interest, making it hard to make informed privacy choices. We argue that the process for seeking consent for the collection of sensor data should communicate not only how inferences were generated and what they are, but also specifically how they are to be used.

1 Introduction

Adequate privacy depends on the ability of end users to give or withhold free and informed consent to *proposed uses* of data [7]. However, digital service companies are collecting massive amounts of data about people who use their platforms, and using the data to make inferences about users' interests, life circumstances, and personal characteristics [9]. Inferences are new data that ca be derived by processing existing data about someone or something. Despite regulations such as the European Union's General Data Protection Regulation (GDPR) which requires companies to disclose information to end users about data collection and processing by the platforms they operate, people are often given very little information about the inferences platforms make about them.

Modern cars are digital platforms equipped with a large array of sensors that collect many different types of data which can be particularly invasive [4]. Applications are making and using inferences based on driving data (e.g. monthly grade 'B' on a driver's report card[1], safe or unsafe driver) to convey to users aspects of the data that have been collected and processed about them, with the assumption that people understand and consent to how their data are used to generate the inferences. However, most people have limited or no understanding of the kinds of data recorded by their cars, let alone the ways those data could be used [8].

Privacy in sensor-based systems must be studied in the context of actual lived experiences, because everyday behavior may differ from stated preferences [2], and the context in which users experience being sensed contributes to their privacy perceptions [3]. In this study, we explored participants' reactions to inferences we generated from data about their actual driving behavior, collected over a 3-month period. Our preliminary findings show that to make sense of inferences, most participants rationalized them, making the inferences fit within their positive image of their own driving. The meaning attributed to the inferences was derived from a combination of the inference shown and how participants interpreted it in the context of their beliefs about themselves. Therefore, it may be difficult for individuals to conceptualize how a platform might interpret and use inferences about them, particularly if an inference presents them in negatively from the platform's perspective. This suggests that to achieve transparency about inferences and meaningful consent to the collection and use of data, platforms should provide end users with information not only about why an inference was made but also how it will be interpreted and used.

USENIX Symposium on Usable Privacy and Security (SOUPS) 2021.
August 8–10, 2021, Vancouver, B.C., Canada.

[1] https://www.motosafety.com/features.html

2 Related Work

While people care about their privacy, they feel they have little awareness of and control over the collection and use of their data [1]. Das et al. argue that "privacy is about giving people meaningful choices when it comes to collection and use of their data and about giving them sufficient details about these choices to make informed decisions" [1]. Studies have investigated transparency in data collection and inferences.

Rader and Slaker [6] interviewed users of activity trackers to understand folk theories about data collected by the trackers. They found that users identified relationships among data types by noticing when numbers changed simultaneously in the interface. However, participants did not understand that what they saw were inferences. Rader et al. [5] investigated user reactions and perceptions of inferences created about them by Facebook and Google. They found that the framing of participants' understanding was limited to things users were directly knowledgeable about, e.g. past behaviors, self-perceived interests, and perceived interests of family members. Weinshel et al. studied how visualizing longitudinal and inference-level information about online tracking impacts user knowledge, perceptions and attitudes [9]. They found that the visualizations provided users with a richer visibility into kinds of information collected, but also what can be inferred from the data. The authors argue that this can help users make more informed privacy decisions and policy demands.

While there have been efforts to investigate transparency about inferences, the work primarily focused on inferences based on web/online tracking [5, 9]. In addition, the findings described above focus most heavily on user awareness of the existence of data collection and inferences [6], but not necessarily how users understand the meaning and potential consequences of the inferences. To improve transparency and give users more control over their data, we need to understand how people currently make sense of how their data are processed and used.

3 Method

3.1 Procedure

The study took place during October 2020-March 2021. To begin the data collection, each participant installed an OBD-II adapter by plugging it into their car's On Board Diagnostics port, which is the same port used by car repair technicians to access data from the car's computer. The adapters, from Zubie[2], recorded information about: car make and model, speed, braking and acceleration, fuel type and use, beginning and end times of trips, and geolocation of the device.

After 12 weeks of data collection, researchers generated a report for each participant consisting of visualizations of data collected as a result of their driving activities. An *individual* version of the report was created for 30% of participants, and an *aggregated* version was created for 70%. The *individual* reports included only each participant's own driving data, while the *aggregated* reports included individual driving data, comparisons with other participants, and data from third party APIs. Report type was randomly assigned to each participant.

During an approximately 90-minute video interview, participants were shown their driving report and asked to think out loud while reacting to its contents. The interviewer then asked follow-up questions about things that stood out to participants and why, and probed for beliefs about how the information in the report was generated, where it came from, and how it might be used.

3.2 Driving Style Inference

The *aggregated* version of the report included an inference characterizing the participant's driving style as *defensive*, *normal*, or *assertive*. For each category, we calculated the percent rank of each participant compared with all other participants for three different data types: the number of hard brakes, rapid accelerations, and speeding events. The percent ranks were summed, and the resulting variable was divided into three quantiles. The lowest quantile participants were designated as *defensive* drivers, the middle quantile as *normal*, and the highest quantile as *assertive*. These labels were chosen as to sound familiar but non-offensive (e.g., assertive instead of aggressive). In addition to the inference, the report presented visualizations of the braking, accelerating and speeding data, but did not explain the driving style calculation or offer a qualitative interpretation of the inference category. Participants were simply informed of their category label, without being told what the other categories were. This was a design choice intended to imitate the way that social media platforms present advertising inference categories.

3.3 Participants

Participants were recruited using a university-run subject pool which solicits members from the broader community. Eligible participants had no formal computer-related training, were at least 18 years old, had access to a car for the 12-week study period that was compatible with the OBD-II adapter, drove at least 4 days a week, and lived in the Lansing/East Lansing area. University students and close contacts of research team members were ineligible, as were people who expected that individuals under the age of 18 might drive their car during the study. Eligible participants were selected for the study prioritizing age diversity and gender balance.

Each participant met with a member of the research team by video before consenting to the study to learn about the study procedures, what data would be collected, and how to pause data collection. Consent was obtained both from the participant and from other people over the age of 18 who would drive the participant's car during the study. After consenting, participants were mailed an adapter and installation instructions. Participants received a $5 Amazon gift card for the initial meeting, plus a lump sum (Amazon gift card or

[2] https://zubie.com/

Zelle transfer) at the end of the study including $10 for each week of data collection plus $20 for the interview and $5 for completing a final survey ($150 total). Forty participants completed the 12-week data collection; 1 did not complete the interview and final survey. This study was approved by our institution's IRB.

3.4 Analysis

This paper focuses on preliminary analyses of interviews of the 28 participants who received the *aggregated* report (16 women, 12 men; mean age: 45) which included the driving style inference. We conducted an iterative inductive qualitative analysis in which four members of the research team did initial coding for themes in 4 of the 28 interview transcripts. The team met to discuss the themes that were emerging, developed a codebook, and then separately coded another 4 transcripts. After another discussion, the codes were revised and all 28 transcripts were divided among the team members and coded using the final codes. These codes focused on participants' reactions to the driving style inference, and how they framed their understanding of the inference in relation to their beliefs about themselves as drivers.

4 Findings

Our analysis identified patterns in the kinds of information that participants used to rationalize the inference about their driving behavior, which we refer to as frames of reference ('frames'). Participants used frames not only to interpret how an inference was produced, but also to assess whether the inference applied to them and what they believed it said about them as a driver. Though some participants expressed feelings of apprehension and dread ("embarrassing," "scary") about their driving style, they appeared to be relieved when their inference was not as negative as they expected it to be. Generally, the majority of participants affirmed that the inference they were assigned was an accurate representation of their driving. For example, this was P29's initial reaction to the driving style inference:

> Assertive driver, I can imagine that. [...] I mean, when I'm driving, I want to get from point A to point B quicker than slower I guess. So try to time it out. *–P29, man, 31 y.o., assertive*

Participants who speculated that the driving style inference was based on some combination of hard brakes, speeding, and accelerations typically identified specific data points from the report that allowed them to make sense of the inference. For instance, P06 (woman, 30 y.o., normal) commented "Wow, normal's weird" when later noting the amount of hard brakes and speeding that appeared on her report. She used this as evidence to support her assertion that she "drives like a jerk", contradicting the *normal* inference assigned to her.

Overall, regardless of the inference they were assigned, participants appeared to interpret it to mean something positive about their behavior or attitudes as a driver. We identified four different frames that participants used to rationalize the driving style inference and relate it to their impressions of themselves as drivers.

4.1 Moral Correctness

Nine participants rationalized the inference by explicitly or implicitly relating their driving behavior to some standard of what was good or bad. For instance, some participants associated *defensive* with driving safely, avoiding endangering others, and not engaging in "reckless" driving. Other participants related *assertive* to mean taking quick, decisive action, being proactive in avoiding potential accidents, and "not being an asshole". One participant (P02, woman, 28 y.o.) found it "weird" to be labeled as *normal*, given that she had been previously told she was stressful to drive with. This highlighted her belief that while her driving behavior could be perceived negatively by others, being labeled *normal* itself was positive.

> So [defensive] means, for example, it says that hard brakes events have been very frequent. So I'm trying to, I've been trying to protect myself. I guess so, I don't know... So, like trying to be on the safer side? *–P13, woman, 30 y.o., defensive*

P13 used the data from the report to give further context to her assumptions about the meaning of *defensive.* She reasoned that the number of hard brakes in her report might be an explanatory factor for why her driving was categorized as *defensive*, which was a label she associated with safe driving. The moral correctness frame allowed participants to interpret the inference to imply how good their driving behavior was.

4.2 Social Comparisons

Eight participants used social comparisons to make sense of their driving style inference, evaluating their behavior against that of other drivers. In some cases, considering how their data compared to others caused them to re-evaluate the meaning of the inference. Some of these comparisons were specifically based on data presented in the participants' reports, such as information about how their hard brakes and rapid accelerations compared to other participants, while others were taken from the participants' personal observations and experiences. Here, P16 made an assumption that the inference was based on one's speeding behavior:

> There should be another level up than the assertive, I think. There can be a level. [...] I think I drive assertive but almost all the people that drive fast and I've seen people driving more faster than me like it's normal. *–P16, man, 31 y.o., assertive*

As he recalled seeing others drive faster than him, P16 concluded that the *assertive* label must not encompass particularly negative behavior. That is, since he did not perceive himself to be much of a speeder compared to others, he believed the inference he was assigned was based on a ranking that categorized drivers by the degree of speeding they engaged in. This suggests that by thinking about how their driving data

might be evaluated as better or worse in comparison to others, participants inferred that the inference connoted some form of evaluation and categorization.

4.3 Norms of Driving

Four participants related the inference assigned to them to norms of driving: behaviors or attitudes which they believed to be accepted or performed by most other people. Even when the label *assertive* was acknowledged to have negative connotations, some participants interpreted it to reflect that they were adhering to a standard of behavior appropriate to a given context, such as driving in an area where speeding is common, or where one is surrounded by aggressive drivers. For example, P35 attributed his driving style to the prevailing attitude about driving in the areas he has lived in, and that operating in an environment with aggressive drivers requires one to learn how to be aggressive as well:

> So I've driven a lot in Boston and DC, and they're aggressive drivers. You got to make your way through, so I learned that. – *P35, man, 68 y.o., assertive*

P35 accepted the label of *assertive* because he connected it to "aggressive." He felt this word described his driving, and did not consider it to be a negative quality because of his familiarity with places where aggressive driving would be considered the norm. This frame highlights how an inference could convey information about a participant conforming (or not conforming) to particular accepted driving practices.

4.4 Individual Attributes

Eight participants related the inference to attributes or circumstances, specific to themselves, which explained or were related to their driving behavior. This frame encompassed a broad set of factors such as having a personal preference for getting to another place quickly, family members whose behavior or preferences affected how they drove, individual traits such as being naturally alert or cautious, or a change in lifestyle that meant different driving habits. For example, this was how P17 explained her interpretation of what the *defensive* inference meant to her:

> I'm cautious. Yeah, I guess cautious might be a word. I'm aware of my surroundings, and I'm checking out my surroundings. *–P17, woman, 70 y.o., defensive*

Participants used this frame to relate the inference to information they already believed and accepted about themselves. This, in turn, allowed them to agree with the inferences more readily. Because this information was related to individual attributes, the inference was also interpreted to imply a characteristic they possessed or which could be assigned to them.

5 Discussion

Our preliminary analysis shows that the frame a participant used to interpret their driving style inference indicated the types of context and information they used to determine its meaning. The frame participants used to connect inferences to themselves illustrates how they saw the inference as a mostly positive reflection of their driving behavior. Aggregated across participants and taken to an extreme, participant's reactions could be characterized as: 'being assertive/normal/defensive *means* I am a safe [*moral*], typical [*social comparison*] driver that follows the unwritten rules [*norms*], except when I don't, and that's because of something I already know and accept about myself [*individual attributes*].'

Once people become aware of an inference made about them, their understanding of it becomes a reflection of themselves. This self-serving interpretation may not be related to how a platform could actually use the inference or the data the inference is based on. Though informing an end user about an inference may allow them to know that the inference exists, this awareness may be shallow and disconnected from any consequences the inference may have. People's privacy choices regarding a system are based on what they perceive that system to be capable of doing with their information or saying about them. An understanding that is primarily situated in their own experiences and self-image may not help them speculate about potential uses that would allow them to make privacy choices in line with their preferences. Merely being transparent about what an inference *is* may do little to help people understand what an inference might *mean*, particularly to others who might use it.

Our findings, though preliminary, suggest that efforts to raise awareness that focus solely on informing individuals of the existence of data collection and inferences are not enough for people to make informed choices about allowable processing of sensor data about them. When inferences are presented ambiguously, people draw on various sources to ascribe meaning to them. Interpreting inferences predominantly through rationalizations that place emphasis on the positive, self-reinforcing aspects of the inference suggests that people may have difficulty conceptualizing uses or interpretations that run counter to their self-interest. For people to make informed choices, the consent process must do more than convey the existence or definition of a particular inference sufficiently for an end user to make sense of it; it must also be able to adequately communicate specifically what the inference is to be used *for* by the entity collecting and processing it. If self-constructed frames are a way to ascribe positive meanings to an inference, they may also be a way to communicate potential risks and consequences and help people engage and reflect on the inference in new ways, which could in turn affect their beliefs and behavior around their privacy choices. We will continue to analyze our data to further investigate participants' reactions to other types of data and inferences presented in their reports, the nuances in the relationships between our participants' thoughts of where the data came from, and what this means for their awareness of what it says about them and how it could be used.

Acknowledgments

We thank Joe Freedman and Rick Wash for their assistance with developing the software for this project. We also thank Anjali Munasinghe, Steven Cauthen, Irene Kim, and the BIT-Lab @ 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] Anupam Das, Martin Degeling, Daniel Smullen, and Norman Sadeh. Personalized privacy assistants for the internet of things: Providing users with notice and choice. *IEEE Pervasive Computing*, 17(3):35–46, 2018. doi:10.1109/MPRV.2018.03367733.

[2] Giovanni Iachello, Khai N. Truong, Gregory D. Abowd, Gillian R. Hayes, and Molly Stevens. Prototyping and sampling experience to evaluate ubiquitous computing privacy in the real world. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, CHI '06, pages 1009–1018, 2006. URL: https://doi.org/10.1145/1124772.1124923, doi:10.1145/1124772.1124923.

[3] Predrag Klasnja, Sunny Consolvo, Tanzeem Choudhury, Richard Beckwith, and Jeffrey Hightower. Exploring privacy concerns about personal sensing. In Hideyuki Tokuda, Michael Beigl, Adrian Friday, A. J. Bernheim Brush, and Yoshito Tobe, editors, *Pervasive Computing*, pages 176–183, Berlin, Heidelberg, 2009. Springer Berlin Heidelberg.

[4] Sasha Lekach. Amazon's driver monitoring app is an invasive nightmare. *Mashable*, February 2021. URL: https://mashable.com/article/amazon-mentor-delivery-driver-monitoring-app/ [cited May 25, 2021].

[5] 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. USENIX Association, August 2020. URL: https://www.usenix.org/conference/soups2020/presentation/rader.

[6] Emilee Rader and Janine Slaker. The importance of visibility for folk theories of sensor data. In *Thirteenth Symposium on Usable Privacy and Security (SOUPS 2017)*, pages 257–270, Santa Clara, CA, July 2017. USENIX Association. URL: https://www.usenix.org/conference/soups2017/technical-sessions/presentation/rader.

[7] Robert H Sloan and Richard Warner. Beyond notice and choice: Privacy, norms, and consent. *Journal of High Technology Law*, 14:370, 2014.

[8] Olivia Solon. Insecure wheels: Police turn to car data to destroy suspects' alibis. *NBC News*, December 2020. URL: https://www.nbcnews.com/tech/tech-news/snitches-wheels-police-turn-car-data-destroy-suspects-alibis-n1251939 [cited May 25, 2021].

[9] Ben Weinshel, Miranda Wei, Mainack Mondal, Euirim Choi, Shawn Shan, Claire Dolin, Michelle L. Mazurek, and Blase Ur. Oh, the places you've been! user reactions to longitudinal transparency about third-party web tracking and inferencing. New York, NY, USA, 2019. Association for Computing Machinery. URL: https://doi.org/10.1145/3319535.3363200, doi:10.1145/3319535.3363200.