Author's copy. Appeared in: Nørskov, M., Seibt J., Quick O. 2020. Culturally Sustainable Social Robotics—Proceedings of Robophilosophy 2020. Series Frontiers of AI and Its Applications, IOS Press, Amsterdam.

# Discrimination issues in usage-based insurance for traditional and autonomous vehicles

Martim BRANDÃO[a,1]
[a] *King's College London, United Kingdom*

**Abstract.** Vehicle insurance companies have started to offer usage-based policies which track users to estimate premiums. In this paper we argue that usage-based vehicle insurance can lead to indirect discrimination of sensitive personal characteristics of users, have a negative impact in multiple personal freedoms, and contribute to reinforcing existing socio-economic inequalities. We argue that there is an incentive for autonomous vehicles (AVs) to use similar insurance policies, and anticipate new sources of risk that may lead to indirect and structural discrimination. We conclude by analyzing the advantages and disadvantages of alternative insurance policies for AVs: no-fault compensation schemes, technical explainability and fairness, and national funds.

**Keywords.** Insurance, autonomous vehicles, ethics, indirect discrimination, structural discrimination

## 1. Introduction

### *1.1. A trend towards user tracking in insurance*

Recent days have seen a surge in personal tracking devices for multiple purposes. From weight tracking for health purposes [1], to sex life tracking [2], location tracking for child safety, or GPS tracking for car theft recovery, there is a growing number of tracking devices in the market to fit all concerns. The insurance industry has also started to pick up on such innovations and use them within their products. For example, life insurance companies have started to use social media data to set premium rates [3] and fitness tracker data to set health insurance rates [4, 5]. This can be concerning since health insurance companies are secretive about their algorithms [6] and have been found to use flawed risk assessment methods in the past [7].

A trend towards usage-based policies is happening in car insurance, which promises users cheaper premiums if they install dedicated tracking devices in their cars, or use smartphone apps that track GPS, speed and acceleration data. Premiums are based on quantitative "scores" related to distraction, driving smoothness and other factors. Insurance companies are interested in using such surveillance-based technologies to inform premiums both because: 1) they can attract more low-risk and

[1] Corresponding author: Martim Brandão, King's College London, Bush House, 30 Aldwych, London, WC2B 4BG, United Kingdom; e-mail: martim.brandao@kcl.ac.uk.

hence profitable clients into their company [8], and 2) they create incentives for clients to actively reduce risky behavior by knowledge or fear of raise of premiums [9, 10].

In this paper we deal with vehicle insurance, and try to anticipate issues as the industry moves towards (more) autonomous vehicles. Usage-based insurance products go by different names, such as pay-how-you-drive, black-box insurance, telematics insurance, or usage-based insurance. We will use the term "usage-based insurance" in the rest of the paper. Our contributions are:

1. We discuss the way in which usage-based car insurance could discriminate (Section 2) and negatively impact personal freedoms (Section 3),
2. We anticipate the likely changes to the sources of risk in AVs and how these may contribute to further discrimination (Section 4),
3. We discuss the advantages and disadvantages of alternative insurance policies in light of the previous discussions of bias and discrimination (Section 5).

### *1.2. Current criticism of usage-based car insurance*

Before we expand on discrimination, we briefly describe the issues that have been raised regarding usage-based car insurance.

#### *1.2.1. Cream skimming*

Since usage-based insurance further refines and narrows down risk categories, it can lead to larger differences of premiums between the lowest and highest paying categories. This is one of the criticisms of usage-based insurance, called "cream-skimming" [8]. Such price differences may end up excluding high-risk groups from the service. This criticism assumes that risk is accurately predicted, but in general victims of skimming are not "high-risk" groups per se, but those groups that are *predicted to be high-risk*.

#### *1.2.2. Privacy*

Tracking a client's location or health state for risk-assessment purposes comes with privacy violation issues. One problem is if the employees of the insurance company or a data analysis contractor have access to sensitive personal information [11]. Even if data is anonymized, de-anonymization algorithms can be used to recover personal identities, especially with location data [12]. Another problem is the difficulty in building completely secure computer systems in practice. Finally, there are privacy concerns regarding how user data is shared with third parties, for example for advertising [12] or repair and maintenance purposes [13].

#### *1.2.3. Black box nature is profitable and unaccountable*

The algorithms used to arrive at insurance premiums, including those of existing usage-based "black-box" policies, are not disclosed to their customers because of trade secrecy [7]. Thus, car users do not know exactly how they should drive to be considered low risk [10]. One underlying issue is that the usage of black-box algorithms is both profitable and unaccountable, because they can be marketed as accurate and fair, and because the decision process is complicated and proprietary [14]. Additionally, in case the actual decision process is not inherently transparent to both clients and auditing agencies then "[their] designers are not incentivized to be careful in its design [and] performance" [14].

### *1.2.4. Discrimination*

Another criticism is that usage-based policies may discriminate specific social groups, such as older drivers [15], female drivers [16], or people with physical or cognitive disabilities. Even if these drivers passed the appropriate exams required for obtaining or renewing a driving license, the presence of certain driving acceleration or distance profiles could trigger high premiums in ways that correlate with protected personal characteristics of the drivers. Hence, there are potentially groups of people that are more likely to be penalized. This critique is usually focused only on the existence of a differential assuming that risk predictions are accurately defined [17, 9]. Not much has been said about the kinds of discrimination that can take place, what the impact on people's freedoms could be, and whether these would be unfair. These topics will be the focus of our discussions for the rest of the paper.

## 2. How usage-based car insurance can discriminate

Two factors of discrimination that have been missing in the criticism of usage-based insurance are: 1) problems of bias in algorithms, datasets, or algorithm performance; and 2) problems of structural discrimination that could ensue. In particular regarding bias, the use of big data and highly personal data opens the door to discrimination issues that are present in many of present-day machine learning (ML) technologies [18, 19, 20]. ML algorithms typically find correlated proxies that solve a task, but these could be inherently wrong [14], or induce lower accuracy for minority populations due to dataset representation issues [18, 21, 22]. Flaws in risk-computation algorithms can already be spotted in traditional insurance. For example, ProPublica reported 30% higher insurance fees are charged to minority neighborhoods, compared to non-minority neighborhoods of the same risk [6]. These issues can only worsen with the introduction of highly personalized data that correlates with living, working, and socializing locations, as well as physical and cognitive characteristics related to driving behavior. Such data could be used for profiling and redlining of a new form much more pervasive than current ZIP-code policies. This section will further elaborate on the ways in which usage-based insurance can discriminate.

### *2.1. Inaccurate risk metrics*

One typical criticism of risk metrics used for insurance purposes is that they might not be causally related to the actual occurrence of accidents, but be only irrelevant correlations. This could lead, for example, to drivers adopting potentially risky behavior on purpose in order to achieve low scores for premium purposes [9]. For what concerns us here, inaccurate metrics might also be proxies for personal characteristics such as age, race [6] or gender [16]. The use of inaccurate metrics could be seen as a form of procedural unfairness since it originates from incorrect processes being in place to estimate risk. These processes might also lead to outcomes that are biased against people of those protected characteristics, or otherwise arbitrary characteristics outside of people's control. The use of wrong inaccurate metrics can also, in that sense, lead to issues of distributive justice: because premiums can be distributed unfairly across good- and bad-weather states, or male and female drivers, for example, even if these are not true causes of risk.

### *2.2. Indirect discrimination through location tracking and redlining*

Even if risk is accurately predicted by geographical location of a vehicle, there could be a problem with usage-based insurance in terms of the distribution of its impact, i.e. in terms of distributive fairness. Data regarding the location of a vehicle can be predictive of risks of theft and vandalism (e.g. using area statistics such as crime rate) and risks of accident (e.g. using traffic, road condition, or historical accident data). Using such data for insurance purposes is similar to redlining practices in the US that use residence-postcode for setting insurance premiums—a practice heavily criticized for being discriminatory [6]. User tracking further refines this kind of policy by using knowledge of, not only the area of residence, but also where a vehicle travels to and travels through in each trip or on average. Such surveillance-based policies can further introduce issues of discrimination. Where passengers usually go to depends on their social circle, where their family and friends live, where they go to work, etc. To pay a low premium, passengers must not only live in a "safe" area, but also work and travel only to "safe" areas. One of the issues here is that these are usually factors that are out of a person's control. At the very least, whether or not a passenger visits her family or friends should not depend on the potential increase of insurance premium. Additionally, location-based measures of risk, such as crime rates, are correlated with characteristics such as race and socio-economic background, and therefore the use of location tracking for insurance has a potential for indirect discrimination on such protected characteristics. Furthermore, the fact that location-tracking-based policies can make transport use more costly to those of lower income, because of how income is correlated with crime-rate, is potentially troubling. Such schemes could introduce or reinforce social inequalities, similarly to previous cases of unintentionally discriminatory urban planning practices [23]. One could also argue that pricing should not be correlated with area crime-rate itself, as this would lead to double punishment of users—of having to pay extra on top of already having to go through the exposure to high crime risk.

### *2.3. Networked discrimination*

One concept relevant to the issues of tracking in car insurance is that of networked discrimination [24], which is a new form of discrimination associated not with personal characteristics but personal networks. The idea is that because of the use of large amounts of data about people and their networks, decisions about people can now be related not only to their personal characteristics, such as race or gender, but also to "who they know". In the context of usage-based car insurance, the nuance is that personal networks are not those represented by online behavior, but those defined by where people physically travel to. If insurance premiums are based on travel statistics, then they will also be related to where the user's personal network lives and travels. As argued by [24], this sort of discrimination is not well modeled by current ethical and legal frameworks.

## 3. Impact of usage-based car insurance on freedoms and society structure

### *3.1. Personal freedoms*

Sophia Moreau defends in [25] that discrimination related to a personal characteristic affects at least three kinds of personal freedoms of those discriminated: negative freedom, domination freedom, and deliberative freedom. We now discuss in which way each of these freedoms is affected by usage-based insurance in particular.

*Negative freedoms.* "Negative freedoms" are freedoms not to have options removed [25]. In the context of usage-based car insurance (autonomous or not), discrimination can make an existing transportation option more expensive (e.g. to older people [15]), and even remove a transportation option entirely if costs are prohibitive (e.g. due to "cream skimming"). Such impact could be acceptable in case other transportation options are made available to vulnerable users, but we should be skeptical of this given the fact that many old and minority people in the modern world live in isolated communities.

*Domination freedom.* "Freedom from domination" as described by Moreau is the freedom from the possibility of future arbitrary control [26]. In our context, car users may come under the domination of insurance company algorithms, whether because they are not sure about the consequences of their travel behavior for future premiums, or they cannot control them. In the paradigm of usage-based policies, any small event (e.g. friend visit, mood change in the case of driving assist, weather change) comes with a possibility of a change of premium.

*Deliberative freedom.* "Deliberative freedom" is the freedom to not be reminded of personal characteristics (such as race or religion) [25]. In our context, deliberations are expanded to those of the person's social circle (e.g. "my friends live in a black neighborhood", "my son goes to school in an unsafe area"). Through such insurance policies, car users will keep being reminded of the socio-economic background of their family and relatives, about aspects of their social circle, about the riskiness of their place of residence, work, leisure, etc. This may also have an important impact in the spread and reinforcement of social prejudice and structural discrimination. Note that a violation of deliberative freedom is separate from, and may happen in addition to, the unfairness of pricing differences related to personal characteristics—and the associated negative freedom violation.

### *3.2. Structural discrimination*

At the systemic level (i.e. structural discrimination [27]), the introduction of a new fine-grained dimension of risk to transportation can reinforce existing socio-economic prejudices. As studied in the literature of environmental discrimination, inequalities in access to transportation can lead to reinforcing inequalities of access to job opportunities, housing and infrastructure quality, and social isolation [23]. If predictors of accident risk are related to road quality, often correlated with socio-economic characteristics of residents, then there will be an incentive for cars not to travel to these areas—further contributing to exclusion and the reinforcement of prejudices related to these areas. The same can be said about risks of vandalism and theft.

New subjects of systemic discrimination could also form: if risk can be predicted by new factors, let us say weather for the sake of example, then areas with characteristics problematic for AVs (foggy or rainy) could become more expensive to travel to, having a deteriorating effect in the local economy and isolating the area in terms of transportation. This, in turn, could lower investment in infrastructure, lower housing prices, and attract low-income residents thereby creating a spiral of risk (as per insurance assessment) and socio-economic reconfigurations.

## 4. Extension to autonomous vehicles

We now turn to a discussion of how the previous insights regarding discrimination may apply to autonomous vehicle insurance.

### *4.1. How AVs may further favor tracking*

Car insurance products typically cover costs related to driver-caused accidents, no-fault collisions, or theft and vandalism. AVs are expected to be safer than driven vehicles, and thus shift risks from human-caused accidents to collisions and theft. This shift to causes that are not related to drivers' "proper use" of the vehicle also motivates a shift towards predicting risk from other metrics not related to the driver's personal characteristics. For example, both collisions and theft risk can be related to location data due to road quality, traffic, crime rate, and other area statistics that we will discuss. There is also evidence of interest from insurance companies in exploring different regimes of insurance for AVs. For example, the Innovate UK-funded project DRIVEN [28] is investigating the use of new risk profiling tools for AV insurance using data collected from each car. Possible factors for favoring usage-based insurance policies in AVs include:

*Cost.* Because of their increased cost, AVs might be more appealing targets of theft and vandalism, in which case car data could be used to obtain a proxy for risk (whether the location "looks" likely of theft/vandalism given visual or location data).

*More sensors for free.* AVs need to be equipped with a multitude of sensors of various types, such as cameras (for visually detecting objects), lasers (for measuring distances to objects), GPS (for location), odometry systems (for estimating vehicle state and location), etc. Such sensors are a crucial requirement of autonomy. Once vehicles have such sensors, however, the door is open to other uses of the data. Some concerns exist, for example, regarding the privacy issues associated with having a large amount of such sensors capturing data of passersby. In regard to insurance, such sensors open the door to methods which try to predict risk "as you go". For example, the risk of accident could be predicted from measurements of road condition, kinds and amounts of surrounding vehicles and pedestrians, historical accident data of the current location, etc. Similarly, the existence of visual and location data opens the door to predicting theft and vandalism from statistics of where the vehicle is parked.

*New predictors of risk.* Because the nature and processes guiding AV driving are different from traditional cars, there will be new predictors of risk that do not exist currently. For example, AVs depend on the performance of their pedestrian detection, traffic sign detection, other-vehicle detection, state estimation, etc.—and this performance influences risk. As we will discuss next, the performance of AVs might also directly depend on weather, lighting, road, or other conditions, in new ways compared to traditional cars. Since most of the risk of AVs is likely to shift towards such sensor- and decision-related factors, there is an interest in the insurance industry using predictions of such kind of risk—which require car data.

### *4.2. A new source of indirect discrimination: biased performance*

As we have just discussed, a shift to AVs can introduce new predictors of risk and an incentive to apply usage-based insurance policies. We now elaborate on a new source of risk in AVs that can lead to indirect discrimination—algorithm performance differentials.

#### *4.2.1. Emotion, responsiveness and tiredness recognition*

For semi-autonomous vehicles which require drivers to stay alert and intervene when necessary, estimations of the driver's ability to quickly respond are predictors of risk. Estimations of driver depression, anxiety, tiredness, or road awareness could be used by algorithms through cameras and other sensors tracking the driver's state. One of the problems here is that the performance of such algorithms could vary with respect to some arbitrary characteristics of the drivers. For example, the performance of facial analysis has been shown to be biased (have lower accuracy for black females) [18], and similarly for other tasks such as emotion recognition [29]. Even if one ignores the intrinsic issues with basing premiums on such factors, the biased-prediction problem is that some people could be more likely categorized as "unawake", "depressed", etc., when in fact they are not.

#### *4.2.2. Pedestrian detection*

Autonomous vehicles rely on detecting pedestrians and predicting their future motion in order to produce safe driving trajectories. Recent research has shown that state-of-the-art image-based pedestrian detection algorithms also have biased performance: specifically they are more likely not to detect children [19] and people of darker skin tones [21]. As discussed in [19], the reasons for this bias can stem from multiple factors: from training set distributions, to algorithmic issues related to contrast or distinguishing noise from small objects. While similar analyses have not been done on LIDAR-based detection algorithms (which are more popular in AVs), it is reasonable to assume that there could be a similar performance bias, even if just related to the size of pedestrians. However, image-based pedestrian detection will likely be used together with LIDAR in order to differentiate between people, animals, and harmless objects such as rubbish. Again, the issue of biased performance is likely to appear. One of the problems with this bias is that specific social groups may be more likely victims of accidents with AVs (i.e. those less likely to be detected on the road). Additionally, and in relation to insurance, if such bias exists then the risk of driving can be associated with the social makeup of the geographical areas the car travels through. A car that often drives near schools or parks is more likely to find children and animals, and therefore be subject to higher risk of accident. Premiums could then be biased, as a result, towards people who frequently drive through such areas.

#### *4.2.3. Traffic sign detection*

For similar reasons to the above, traffic sign detection performance is likely to also be biased. For example, if algorithms are trained in roads where traffic signs are different from the deployed location, then detectors might not detect traffic signs correctly. Therefore, other predictors of risk in AVs could come to be the local appearance, cleanliness, and vandalism of traffic signs; and whether traffic signs are mapped or have to be detected from image-data. Again, these could relate to geographical location in meaningful ways or even to the social characteristics of the locations (e.g. road infrastructure quality is often related with socio-economic indicators).

#### *4.2.4. Vehicle state estimation and prediction*

AVs rely on estimations of the speed, acceleration and intentions of other vehicles on the road in order to compute safe driving trajectories. Again, the performance of such estimations and predictions could depend on factors such as vehicle type, size, visual characteristics, road conditions, or vehicle speed itself. They could also vary with the driving style of the other vehicles or vehicle drivers. Therefore there is a problem of basing risk predictions for premium computation on geographical location, if that correlates with the kind of vehicle that is most likely to be present (e.g. highways vs dirt roads, high vs low income areas).

#### *4.2.5. Weather and illumination*

Weather and time of day can have a devastating impact in image-based algorithms such as pedestrian and vehicle detection. While much work is currently being done in addressing these issues [30, 31], it is likely that performance will not be equal over all weather and light conditions. Therefore, driving risk could be estimated from geographical location and local weather, time of day, and road lighting conditions. Particularly quality of road lighting could be correlated with socio-economic indicators of the areas involved.

### *4.3. Further domination*

The added complexity in the way risk can be predicted in AVs can lead to further domination of people’s decisions. Additionally, the intricate relationships between these predictions and socio-economic factors could contribute to the reinforcement (or introduction) of indirect discrimination issues. The new risk-prediction factors described above could lead to algorithmic bias of a greater extent than what is present in traditional insurance and usage-based insurance of traditional cars. Furthermore, if social groups that are more likely to be considered high-risk, such as minority or high-crime-rate residents and their family members, lose access to AVs, it might reinforce or exacerbate current structural discrimination reflected in inequalities of opportunities. Importantly, even if AVs do not become major transportation options, and do not cause the reduction of quality of public transport, biased access might defeat one of the purposes of AVs—that is to widen the access to safe transportation. Note that the same discrimination issues we have been discussing apply to AVs whether they are privately owned, or autonomous taxis and ridesharing. If these services are allowed to fluctuate their fees based on the locations that the passengers want to go to (and the associated likelihood of crime, or pedestrian/sign/vehicle misdetections) then it is part of the services’ interests to use similar methods of usage-based risk prediction.

## 5. Solutions

In light of the previous discussions of discrimination, we will now turn to the potential advantages and disadvantages of various AV insurance schemes.

### *5.1. Usage-based no-fault compensation scheme*

No-fault compensation schemes are policies where the insurer pays without need to prove liability (i.e. without needing to prove whose fault the accident is). This kind of solution is potentially cheap, because there is no need to pay for legal and inspection labor. This option also has the potential to be transparent, since insurance pricing can be associated with a vehicle model (given risk statistics collected on that model) and the buyer can be informed of this risk at purchase time [32]. This kind of solution could have a positive market influence by incentivizing manufacturers to build safer vehicles, especially if insurance is provided by the manufacturers' themselves [33]. The issue that concerns us here is that, as we have seen, because risk will not only depend on vehicle model but also on its use (e.g. where it will usually drive through) there is an incentive to use usage-based predictions of risk that fluctuate prices as you drive—that lead to the issues of indirect and structural discrimination described in previous sections.

### *5.2. Discard usage-based insurance on no-fault schemes*

A more drastic solution is to use no-fault compensation schemes where usage-based pricing is explicitly forbidden. While problems of explicit usage-based personalization of premiums are removed in this case, we note that there is still a chance for indirect discrimination to leak through. At first sight, each vehicle model is associated with an annual insurance premium tied only to the model and not to its usage or user. However, given that different models are likely to be used by different people—depending on the model's price, appearance, age, etc.—the risk statistics collected on each vehicle are also likely to correlate with the socio-economic discrimination factors we have mentioned. In other words, in no-fault compensation schemes the premium of a vehicle model can still be related to who typically owns the model and where they typically travel to.

### *5.3. Explainability and auditing*

Another potential solution to the discrimination and domination issues described is for insurance pricing algorithms of usage-based (e.g. no-fault) policies to be explainable. This means that both vehicle users and auditing agencies can inquire the algorithms for actionable reasons behind the pricing. If explanations were provided with clear guidance, then users would be able to know exactly what to change about their behavior in order to lower premiums. Additionally, requiring insurance algorithms to answer contrastive questions such as "what would change about my policy if I lived in area X instead of Y?", "where should my most-visited places be moved to in order for my premium to be as low as possible?", then users would be empowered to build discrimination cases against insurers, or to decide whether to switch insurance or car provider. Similarly, auditing agencies could use such queries on a dataset of (virtual or real) users to examine whether companies were indirectly discriminating certain groups. Auditing agencies would similarly be responsible for ensuring that explanations are actionable and clear enough to guide user behavior and auditing actions.

This solution is highly technical and suffers from issues of potential gaming by both insurance companies and users. For example, insurance companies would have incentives to implement algorithms that use deceiving explanations [34] or optimize for the "fairness" metrics used by agencies—thus potentially ignoring other important aspects of the socio-technical problem that are not coded in auditing objectives [35, 36]. Insurance companies could also resist this option due to the possibility of users trying to "game" the system in order receive lower premiums. On the other hand, the user empowerment provided by explanations could have a positive influence on the market, in terms of insurance providers being interested in providing pricing schemes that are acceptable to their users.

### *5.4. Fairness constraints and auditing*

In line with recent efforts in the machine learning community, another attempt at a technical solution that can be made is to algorithmically enforce fairness at the level of usage-based pricing algorithms. In this option, regulation could force companies to satisfy certain formal definitions of distributive fairness, such as satisfying a maximum price difference between highest- and lowest-risk groups, a maximum insurance cost, independence between cost and living area or weather, etc. Auditing agencies could then verify whether these definitions were being respected.

Again, as a technical solution this has the potential to be gamed by insurance providers, and it is also likely to suffer from inappropriate consideration of all relevant aspects of fairness [35, 36]. Finally, given the potential ramifications of such complicated technical solution, in terms of new nuances of discrimination and impact, it is not clear how such an approach would be better for society than purely discarding usage-based insurance altogether.

### *5.5. National or state-level fund*

The last option we consider here is a national or state-level fund. Given the discussions in this paper, this is in our view the only solution where the discrimination-based freedom-violation issues can be overcome. In this case, everyone can supposedly afford premiums since they are spread across a large pool of tax payers, and so negative freedoms are not violated. Premiums are simple, hence there is not a problem of domination freedom. And finally, they are "equal" (or proportional to income) and hence would not reinforce or exacerbate existing structural inequalities. A counter-argument can be made that "low-risk" (or "high-income") users would be discriminated because they would be paying more than they should. A possible reply to this is that such discrimination would be acceptable since the taxes would be such that everyone could afford them, and hence negative freedoms would not be violated. The impact on high-income users would similarly not be significant enough to lower the amount of options that are available to them.

## 6. Conclusion

In this paper we discussed the ways in which usage-based insurance can discriminate and impact personal freedoms, both for traditional vehicles and their newer autonomous forms. We discussed how indirect discrimination can arise from inaccurate risk metrics, indirect location redlining, and a new form of networked discrimination related to the spatial configuration of a person's social circle. Then, we analyzed how these kinds of discrimination can affect negative, domination and deliberative freedoms. We hinted at new sources of risk in AVs, in particular that of biased performance of their algorithms. We concluded with an analysis of the pros and cons of alternative insurance policies for AVs. Importantly, our discussions show that in what indirect and structural discrimination is concerned, no-fault compensation schemes would not be appropriate solutions to AV insurance—whether usage-based policies are used or not. While some technical solutions, such as using formally "explainable" or "fair" algorithms, could alleviate issues of discrimination and freedom, national and state-level funds would tackle discrimination issues at a more fundamental level.

One limitation of our paper is that the anticipation of risk predictors in AVs only suggests kinds of predictors, but not how they might quantitatively affect premium distributions in practice. It would be interesting to use quantitative models of demographics, algorithm performance differentials and other factors to predict the distribution of insurance premiums that would arise. We believe such quantitative work could also help inform anticipatory policy.

## References

[1] K. Crawford, J. Lingel and T. Karppi, Our metrics, ourselves: A hundred years of self-tracking from the weight scale to the wrist wearable device, European Journal of Cultural Studies 18(4–5) (2015), 479–496.
[2] K.E. Levy, Intimate surveillance, Idaho L. Rev. 51 (2014), 679.
[3] L. Scism, New York Insurers Can Evaluate Your Social Media Use - If They Can Prove Why It's Needed, Wall Street Journal, 2019. https://www.wsj.com/articles/new-york-insurers-can-evaluate-your-socialmedia-useif-they-can-prove-why-its-needed-11548856802.
[4] S. Jeong, Insurers Want to Know How Many Steps You Took Today, The New York Times, 2019. https://www.nytimes.com/2019/04/10/opinion/insurance-ai.html.
[5] P. Olson, Wearable Tech is Plugging into Health Insurance, Forbes, 2014. https://www.forbes.com/sites/parmyolson/2014/06/19/wearable-tech-health-insurance/.
[6] L.K. Julia Angwin Jeff Larson and S. Mattu, Minority Neighborhoods Pay Higher Car Insurance Premiums Than White Areas With the Same Risk, ProPublica, 2017. https://www.propublica.org/article/minorityneighborhoods-higher-car-insurance-premiums-white-areas-same-risk.
[7] J. Stanley, Pitfalls of Artificial Intelligence Decision-making Highlighted In Idaho ACLU Case, ACLU, 2017. https://www.aclu.org/blog/privacy-technology/pitfalls-artificial-intelligence-decisionmakinghighlighted-idaho-aclu-case.
[8] D.A. Cather, Cream skimming: innovations in insurance risk classification and adverse selection, Risk Management and Insurance Review 21(2) (2018), 335–366.
[9] M. Marabelli, S. Hansen, S. Newell and C. Frigerio, The light and dark side of the black box: Sensor-based technology in the automotive industry, Communications of the Association for Information Systems 40(1) (2017), 16.
[10] E. Zorn, Insurance app has made me a better but more frustrated driver, Chicago Tribune, 2018. https://www.chicagotribune.com/columns/eric-zorn/ct-perspec-zorn-insurance-smartphonemonitoring-app-0429-20180427-story.html.

[11] M. Day, G. Turner and N. Drozdiak, Amazon workers are listening to what you tell Alexa, Bloomberg, 2019. https://www.bloomberg.com/news/articles/2019-04-10/is-anyone-listening-to-you-onalexa-a-global-team-reviews-audio.
[12] A. Taeihagh and H.S.M. Lim, Governing autonomous vehicles: emerging responses for safety, liability, privacy, cybersecurity, and industry risks, Transport reviews 39(1) (2019), 103–128.
[13] B. Hanvey, Your Car Knows When You Lose Weight, The New York Times, 2019. https://www.nytimes.com/2019/05/20/opinion/car-repair-data-privacy.html.
[14] C. Rudin, Stop explaining black box machine learning models for high stakes decisions and use interpretable models instead, Nature Machine Intelligence 1(5) (2019), 206–215.
[15] M. Punjwani, How does black-box car insurance work?, MoneySuperMarket, 2019. https://www.moneysupermarket.com/car-insurance/how-does-black-box-insurance-work/.
[16] M. Ayuso, M. Guillen and A.M. Perez-Marin, Telematics and gender discrimination: some usage-based evidence on whether men's risk of accidents differs from women's, Risks 4(2) (2016), 10.
[17] M. Yvell and E. Axelsson, Implementation of Usage-Based Insurance solutions: A qualitative analysis of a technology-based insurance model from the perspective of the Swedish insurance industry, Master's thesis, KTH Royal Institute of Technology, 2019.
[18] J. Buolamwini and T. Gebru, Gender Shades: Intersectional Accuracy Disparities in Commercial Gender Classification, in Proceedings of the 1st Conference on Fairness, Accountability and Transparency, 2018, 77–91.
[19] M. Brandao, Age and gender bias in pedestrian detection algorithms, in Workshop on Fairness Accountability Transparency and Ethics in Computer Vision, CVPR, 2019.
[20] J. Angwin, J. Larson, S. Mattu and L. Kirchner, Machine Bias: there's software used across the country to predict future criminals. And its biased against blacks. ProPublica 2016, 2016.
[21] B. Wilson, J. Hoffman and J. Morgenstern, Predictive Inequity in Object Detection, arXiv preprint arXiv:1902.11097 (2019).
[22] B. Taati, S. Zhao, A.B. Ashraf, A. Asgarian, M.E. Browne, K.M. Prkachin, A. Mihailidis and T. Hadjistavropoulos, Algorithmic Bias in Clinical Populations Evaluating and Improving Facial Analysis Technology in Older Adults With Dementia, IEEE Access 7 (2019), 25527–25534.
[23] A. Golub, R.A. Marcantonio and T.W. Sanchez, Race, space, and struggles for mobility: Transportation impacts on African Americans in Oakland and the East Bay, Urban Geography 34(5) (2013), 699–728.
[24] D. Boyd, K. Levy and A. Marwick, The networked nature of algorithmic discrimination, Data and Discrimination: Collected Essays. Open Technology Institute (2014).
[25] S. Moreau, Discrimination and Freedom, in The Routledge Handbook of the Ethics of Discrimination, Routledge, 2017, 164–173.
[26] P. Pettit, A theory of freedom: from the psychology to the politics of agency, Oxford University Press on Demand, 2001.
[27] H.-Y. Liu, Structural Discrimination and Autonomous Vehicles: Immunity Devices, Trump Cards and Crash Optimisation, in Frontiers in Artificial Intelligence and Applications, Vol. Volume 290: What Social Robots Can and Should Do, 2016, 164–173.
[28] DRIVEN, Project website, 2019. https://drivenby.ai/.
[29] A. Howard, C. Zhang and E. Horvitz, Addressing bias in machine learning algorithms: A pilot study on emotion recognition for intelligent systems, in 2017 IEEE Workshop on Advanced Robotics and its Social Impacts (ARSO), IEEE, 2017, 1–7.
[30] F. Yu, W. Xian, Y. Chen, F. Liu, M. Liao, V. Madhavan and T. Darrell, Bdd100k: A diverse driving video database with scalable annotation tooling, arXiv preprint arXiv:1805.04687 (2018).
[31] C. Linegar, W. Churchill and P. Newman, Made to Measure: Bespoke Landmarks for 24-Hour, All-Weather Localisation with a Camera, in Proceedings of the IEEE International Conference on Robotics and Automation (ICRA), Stockholm, Sweden, 2016.
[32] M.A. Geistfeld, A roadmap for autonomous vehicles: State tort liability, automobile insurance, and federal safety regulation, Calif. L. Rev. 105 (2017), 1611.
[33] M. Schellekens, No-fault compensation schemes for self-driving vehicles, Law, Innovation and Technology 10(2) (2018), 314–333.
[34] D. Slack, S. Hilgard, E. Jia, S. Singh and H. Lakkaraju, How can we fool LIME and SHAP? Adversarial Attacks on Post hoc Explanation Methods, arXiv preprint arXiv:1911.02508 (2019).
[35] B. Green and S. Viljoen, Algorithmic Realism: Expanding the Boundaries of Algorithmic Thought, in Proceedings of the ACM Conference on Fairness, Accountability, and Transparency (FAT*), 2020.
[36] R. Overdorf, B. Kulynych, E. Balsa, C. Troncoso and S. Gurses, POTs: Protective optimization technologies, arXiv preprint arXiv:1806.02711 (2018).