# Can Telematics Improve Driving Style? The Use of Behavioral Data in Motor Insurance

Alberto Cevolini [1] , Elena Morotti [2,*] , Elena Esposito [2,3] , Lorenzo Romanelli [4] and Riccardo Tisseur [4] and Cristiano Misani [4]

[1] Department of Education and Human Sciences, University of Modena and Reggio Emilia, 42121 Reggio Emilia, Italy; alberto.cevolini@unimore.it
[2] Department of Political and Social Sciences, University of Bologna, 40125 Bologna, Italy; elena.esposito@uni-bielefeld.de
[3] Faculty of Sociology, Bielefeld University, 33615 Bielefeld, Germany
[4] Swiss Re Europe S.A., Rappresentanza per l'Italia, 20121 Milan, Italy
* Correspondence: elena.morotti4@unibo.it

**Abstract**

Motor insurance can use telematics data not only to understand individual driving style but also to implement innovative coaching strategies that feed back to the drivers, through an app, the aggregated information extracted from the data. The purpose is to encourage an improvement in their driving style. A precondition for this improvement is that drivers are digitally engaged, that is, they interact with the app. This paper proposes a narrow understanding of the term engagement, referring to users' interactions with the app. This interaction is also a behavior producing specific data that can be tracked and used by insurance companies. Based on the empirical investigation of the dataset of a company selling a telematics motor insurance policy, our research investigates if there is a correlation between engagement with the app and improvement of driving style. The analysis distinguishes different groups of users with different driving abilities, and takes into account time differences. Our findings contribute to clarifying the methodological challenges that must be addressed when exploring engagement and coaching effectiveness in proactive insurance policies. We conclude by discussing the possibility and difficulties of tracking and using second-order behavioral data related to policyholder engagement with the app.

**Keywords:** behavioral data; usage-based insurance; engagement; coaching; automotive telematics

## 1. Introduction

Insurance is a fundamental risk-transfer mechanism of modern society. The risks of the insured are financially *transferred* to the insurer and at the same time *transformed*: the damage that would be financially ruinous for an individual is distributed by the insurer among all the members of the same pool and thus becomes sustainable [1–3]. Despite the effectiveness of this risk pooling and spreading mechanism, insurers have an interest in preemptively reducing the risks transferred by the insured. Since many risks depend on individual behavior, as in the illustrative case of motor insurance, the insurer's preventive activity should move from, and act upon, the behavior of the insured. But policyholders' behavior is in principle unobservable.

Since the early 2000s, insurance companies selling third-party liability motor insurance policies have invested heavily in the use of telematics data to track drivers' behavior. The data collected should make it possible to assess policyholders' risk profile and adjust their policy premium accordingly. The insurance industry terms this opportunity usage-based insurance (UBI).

Business experience over the last two decades shows a significant evolution in the use of behavioral data. Insurance companies not only use it to refine the risk profile of policyholders but also feed the aggregated information they obtain from behavioral data back to the drivers. The aim is to promote greater awareness by policyholders of their driving style and to encourage a change in driving behavior in the case that their driving habits show criticalities that increase risk exposure. In this respect, insurance companies speak of *coaching* [4–6]. The usual abstract assumption is that coaching works because policyholders who receive feedback improve their driving style and become better drivers [7] (p. 22).

The effectiveness of coaching, however, is yet to be proven. Our hypothesis is that it depends on the willingness of users to take in and use the information fed back by the insurance company in order to motivate them to change their behavior, which is usually conveyed by a digital app. A rapidly developing strand of research [8–10] calls this kind of activity *engagement*. By engagement, we mean the time and effort that individuals put into improving their risk profile. We propose to distinguish a broad and a narrow sense of engagement. A recent paper [11] exemplifies the use of the term in a broad sense: policyholders who took out health and life insurance based on behavioral data are considered 'engaged' when they participate in pre-established programs including diagnostic screening, gym membership, and daily exercise to promote a healthy lifestyle.

For our research on telematics motor insurance, however, it is important to focus on a narrow sense of the term 'engagement' to refer exclusively to the users' interaction with the app. Our research assumes that this engagement is also a behavior, which like driving behavior is (or can be) tracked by digital devices. If by 'behavior' we do not only mean 'the use of the car' but also 'the use of the app', then insurance companies have to deal with two different types of behavioral data—behavioral data on driving style, and behavioral data on users' interaction with the app.

In this article, we want to explore the interplay of information feedback and behavioral change. According to this approach, the usual UBI formulation should be clarified. The success of proactive strategies implemented by insurance companies is, in fact, based on two different types of behavior that can both be tracked by the telematics app: not only the driving behavior of policyholders but also their interaction with the app, i.e., their being engaged. Insurance companies selling telematics insurance policies collect a lot of data about both behaviors: use of the car and use of the app. This is, in our opinion, a crucial novelty.

Behavioral data are considered a "remarkable advance" in automobile insurance [12] (p. 662). Previously, the insurance industry could only use variables related to fixed characteristics of the policyholder and the vehicle, many of which, such as age and gender, are not causally related to the risk of getting into a crash. They are proxy variables. Behavioral variables, instead, are *causally* related to the risk of road accidents and promise to enable personalized tarification, which may be considered a fairer policy premium setting system [13,14]. Moreover, as we have seen, behavioral data processing can be carried out to implement coaching strategies and possibly improve policyholders' driving behavior. Behavioral data processing, thus, is expected to impact policyholders (who know that their behavior is monitored), insurance companies (which can improve their predicting capacity

by combining behavioral and non-behavioral variables), and the relationship between policyholders and insurance companies (triggering feedback loops).

The objective of our research is to test if the effectiveness of current experiments depends on the integration of these two distinct types of behavioral data. This integration raises a number of new questions: How should engagement be properly defined? How should it be measured? Is there any empirical evidence of a connection between engagement and driving behavior improvement? And how does this connection change over time? To answer these questions, we investigated the dataset of an insurance company selling telematics motor insurance policies. In Section 2, we describe the emergence of the idea of insurance as a loss prevention institution and the evolution of usage-based auto insurance policies over the past two decades. In Section 3, we provide a brief overview of relevant research. In Section 4, we describe the dataset we worked on, the methodology we followed, and some limitations of our study. Section 5 presents our main findings. Section 6 exposes our conclusion and suggests possible directions for future developments of behavioral insurance.

## 2. The Evolution of Usage-Based Auto Insurance Policies

In the mid-1990s, the motor insurance industry began to question the insurance model that merely compensates policyholders' claims. Starting from the assumption that the majority of road accidents are caused by human miscalculations (of driving capability, road conditions, or driving control under certain road and weather conditions), the possibility of insurance companies acting as loss prevention companies began to be discussed. The aim was "stopping claims before they happen" [15] (p. 271). Underpinning this project was the conviction that insurance could not simply be a risk-spreading mechanism. Spreading risks basically means that policyholders transfer their risks to the insurance company, which distributes them over the pool of insured customers. The result is risk mitigation for the customers who feel relieved from the financial consequences of possible future damages.

A consequence, however, can also be that policyholders are less incentivized to take precautionary measures, producing the thorny problem of moral hazard [16,17]. To counter this attitude, it has been suggested to try "to make people more individually accountable for risks" [18] (p. 1). The basic idea was to move from "spreading risks" to "embracing risks": even if policyholders pay for coverage, they should be aware that they retain, at least in part, both a moral and financial responsibility for the consequences of their behavior [18] (p. 3). In the case of auto insurance, this meant that drivers should engage in preventive actions. But prevention first requires an awareness of the risks to be avoided in order for bad driving habits to be removed [15] (p. 278). What remained unclear, however, was how the insurance industry could tackle the problem of bad driving. This is where digital devices used as monitoring devices to produce behavioral data come into play.

The first form of *usage-based insurance* (UBI) tested in the early 2000s was the so-called *pay-as-you-drive* (PAYD) insurance policy [19]. The novelty of this policy was that its pricing system was based on the mileage driven by the policyholders during the policy term. The underlying idea was that mileage is a crucial risk factor statistically related to claim probability. The assumption was that people with low mileage are low-risk motorists and should pay less, whereas people with high mileage are high-risk motorists and should pay more. The PAYD-pricing system was later questioned, as it does not take into account that higher mileage can also mean higher driving experience producing better driving skills. Increasing mileage can be connected with a 'learning effect' [12] that, in turn, might decrease the risk of road accidents. Between young licensed drivers and claim probability, on the other hand, there is a similar statistically significant relationship.

UBI later evolved into *pay-how-you-drive* (PHYD) insurance policies, based on the idea that driving style is causally related to the risk of road accidents and should also be taken into consideration when setting the policy premium. Between statistical variables like gender and age and claim probability, there is actually a strong statistical correlation but no evident causal relationship. Between phone distraction and the likelihood of getting into a crash, instead, there is a causal relationship. PHYD insurance policies, therefore, keep measuring mileage as PAYD policies, but they also track drivers' behavioral characteristics to assess their actual driving style. Drivers' behavior is tracked by means of telemetry packages. PHYD insurance policies usually require the installation of a black box in the car with the policyholder's consent. This black box generates a huge amount of behavioral data that allows the company to monitor the policyholders' driving style—how they steer, how and how often they brake, whether they exceed the speed limit, whether they drive predominantly during the day or at night, and so on. The aggregation of these features makes it possible to assess the individual risk profile and can be used to adjust the policy premium accordingly. This information can also be the basis for coaching services that aim to prevent claims before they occur [20].

The crucial condition to implement coaching strategies is *feedback*. In most advanced telematics insurance solutions, drivers who take out a PHYD insurance policy are supposed to download an app on their smartphone. This app notifies policyholders of the overall score they achieved depending on how well or badly they drove. The same app also communicates the scores achieved in the main features (maneuvers) that the company uses to define the individual driving profile [11,21]. Finally, the app shows every single trip traveled by the insured and indicates exactly whether any criticalities were found and what they are (e.g., where the insured exceeded the speed limit or made a U-turn). This information is made available after driving, not in real time.

By means of feedback, information literally *circulates*, that is, it runs circularly. Drivers disclose information about their driving behavior to the insurance company. The insurance company, in turn, discloses information concerning risk assessment and risk profile to the drivers. Telematics insurance policies, thus, do not simply turn information asymmetry upside down as many scholars argue [22,23]. They rather trigger a circular relationship where behavior produces information, and information is fed back to change behavior. What is really going on in PHYD insurance policies is a kind of 'feedback loop'.

## 3. Previous Research

As shown by a recent bibliometric review of telematics-based auto insurance [24], the literature on telematics motor insurance is very large and ever-expanding. Here, we only focus on the contributions which explore the relationship between information feedback and driving behavior. A recent overview of studies investigating the impact of telematics on road safety points out that there is still scarce research about before/after feedback provision to the drivers [25].

More than twenty years ago, Wouters and Bos [26] (p. 644ff) put forward the hypothesis that drivers who know that they are being monitored might be encouraged to change their behavior, especially if they receive feedback as a result of this monitoring. In their empirical research on a business fleet, Wouters and Bos assessed the effect of this 'behavioral feedback' based on JDR (journey data recorder) by comparing an experimental group with a control group of vehicles for a period of 12 months. A statistically significant accident reduction could be detected only for some of the fleet sets, but the overall accident rate in the experimental group was reduced by 20% after the intervention. However, both monitoring and behavioral feedback were not linked to an insurance policy and lacked

the reinforcement that insurance policies usually provide in addition to feedback, namely, financial incentives.

A decade later, Farmer, Kirley, and McCartt in [27] tested the effects of in-vehicle monitoring on the driving behavior of teenagers, whose crash rates, as is well-known, are consistently higher than any other age group. Feedback, in this case, was notified to their parents on a dedicated website. After 24 weeks monitoring on 85 recently licensed drivers in a suburban Washington DC area, it turned out that there were no statistically relevant changes in driving behavior and that parents themselves made few visits to the website to check the driving behavior of their children. Also in this case, feedback was not associated to an insurance policy.

In 2011, Bolderdijk, Knockaert, Steg, and Verhoef in [28] carried out a field experiment on the effects of a PAYD insurance policy on young drivers' speeding behavior. The basic reasoning was that young drivers are overrepresented in road accidents statistics because they tend to drive at higher speed, and speed is one of the most important behavioral determinants of crash risk. The goal of their research was to test if the provision of financial rewards for keeping the speed limit could encourage young drivers to modify their driving behavior. Participants could check their performance by logging in to a website which provided detailed feedback on speed violations, mileage and night-time driving, and showed by default the prospective overall discount they could earn. The incentive group (ca. 150 participants) showed a modest but significant reduction in speeding, strongly associated to financial incentives (when financial incentives were removed, speeding increased again).

The research that comes closest to the problems we investigate in this article is that of Soleymanian, Weinberg and Zhu [7]. Based on the dataset provided by a major US insurance company offering a PHYD policy, Soleymanian and colleagues were able to observe more than 100,000 customers over a 32-month period. Their main research question was whether there is a statistically significant improvement in the driving behavior of UBI customers compared to usual customers. Their research showed that UBI customers improved their driving score by ca. 9% (from 62.05 in week 1 to 67.87 in week 26), that this improvement was higher in early weeks and for young drivers, and that it did not depend solely on feedback but also on financial incentives. In weeks 11 and 12, 15% of UBI customers dropped out. The consequence was a significant decline in harsh braking, which can be interpreted as the outcome of self-selection: PHYD policies retain the best customers and let bad customers leave.

Soleymanian, Weinberg, and Zhu's research [7] is very important, mainly because it is based on an insurance company dataset and observes UBI customers over time. However, it does not investigate many issues that are crucial for us. For example, if the purpose of coaching strategies is the improvement of policyholders' driving behavior, how should an improvement be properly defined? And how should it be measured by insurance companies that have access to behavioral data? Since improvement should be the result of coaching, how should a coaching process be defined? Are there short-term and long-term coaching effects? Crucial questions for us are also how engagement can be defined and measured, whether there is a connection between engagement and driving behavior improvement, and how this connection changes over time. In our empirical research, we deal with these questions.

## 4. Dataset and Methodology

The data we worked on were taken from PHYD insurance policies based on mobile telematics. In this case, an app in the smartphone replaces the usual black box. Such a replacement has advantages and disadvantages. The smartphone is usually regarded as

an excellent platform for providing users with prompt feedback [29,30]. Moreover, the smartphone detects phone distraction, which is known to be one of the main causes of road accidents and a crucial feature to be integrated into the score evaluation process. The main disadvantage is that telematics data produced by a smartphone are less accurate than the telematics data produced by a black box and require more preprocessing.

During each trip, this app downloads geolocation information from its map provider and records raw data from the GPS, the accelerometer, and the gyroscope as well as from the smartphone system to check for phone usage. The data is processed in real time to assess each driving session based on four key aspects: the attention paid by the driver who should not use the phone while driving the car, compliance with speed limits, cautiousness on the road, and a set of circumstances depending on some external factors (such as driving time during rush hours). Accordingly, four sub-scores are generated to analyze each trip feature separately, namely: 'attentive driving', 'conscious driving', 'smooth driving', and 'contextual' scores. They all range from 0 (poor) to 100 (excellent) and are described in Table 1.

**Table 1.** Description of the four sub-scores assessing each driving session.

| Sub-Score | Description | Inputs |
|---|---|---|
| Attentive driving | It evaluates the driver's level of attentiveness by assessing distractions caused by mobile phone usage. A higher frequency of phone usage events per kilometer results in a lower score for attentive driving. | Count of phone unlock events; total driven kilometres. |
| Conscious driving | It evaluates the driver's adherence to speed limits by assessing both the frequency and magnitude of speed limit violations. A higher percentage of the trip spent driving above speed limits, as well as larger differences between posted speed limits and actual driving speeds when exceeding limits, results in a lower score for conscious driving. | Total driven time; driven time above speed limits; difference between driving speed and speed limit. |
| Smooth driving | It evaluates the driver's level of cautiousness by analyzing the frequency of risky maneuvers during the trip. Risky maneuvers include harsh braking, acceleration, cornering, steering, failure to yield at intersections, aggressive roundabout maneuvers, and U-turns. A higher frequency of risky maneuvers per kilometer results in a lower score for smooth driving. | Count of detected risky maneuvers; total driven kilometers. |
| Contextual | It evaluates the potential impact of external factors on driving behavior, by considering the amount of time spent driving in risky contexts. Risky contexts may include driving during rush hours, at night, or on urban roads. A greater amount of time spent driving in these risky contexts results in a lower score for the contextual sub-score. | Total driven time; driven time during night hours; driven time during rush hours; driven time on urban roads. |

A cumulative trip score is computed as a weighted average of the four sub-scores. It still ranges between 0 and 100, but weights are set by the insurance company, which can balance the importance of the features as it considers most appropriate. Immediately after each driving session, the trip scores are displayed, and all significant critical events are localized on the trip map (as shown by the first screenshot in Figure 1) to ease the scores interpretation. Past trips remain visible and searchable in the app, while a weekly score for the current week is updated on the app homepage as is visible on the right image of Figure 1.
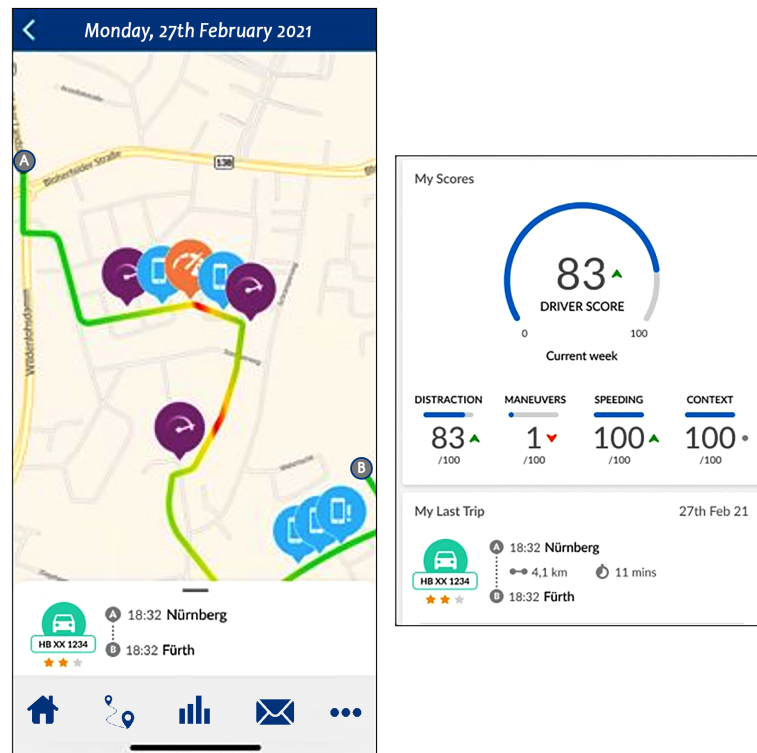
**Figure 1.** On the left, the critical events displayed on each trip map, such as phone distractions (blue icons), risky maneuvers (purple and orange icons), and speeding (highlighted as red road segments). On the right is the app homepage, where the weekly overall driver score and the four sub-scores are updated after each trip.

In many mobile telematics-based insurance policies on the market, the score is used to reward policyholders with financial incentives (e.g., fuel cashback, vouchers, and discount upon renewal of the policy). This is not the case of the UBI motor insurance policy we worked on. On the one side, the absence of financial rewards gave us the opportunity to investigate the pure interplay of information feedback and behavioral change. On the other side, the absence of financial rewards deprived us of the possibility of exploring the relationship between digital engagement and a bonus system based on behavioral improvement. As we explain in Section 4.3, this is an important limitation of our research.

### 4.1. Data Preprocessing

We accessed the trip scores and trip metadata of 498 new customers, onboarded in a 9-month period from March 2022 in a Western European country. The observation period was 35 weeks. In the app weekly summary, the automatically defined week always starts on Monday, disregarding the exact onboarding day of each individual. For this reason, we considered Monday-starting weeks and associated the week indexes $i = 0, 1, \ldots$ to each trip, where $i = 0$ corresponds to the onboarding week for each user, 'aligning' the users according to their own timing. Coherently, for each user $k \in \{1, \ldots, 498\}$, we computed the weekly score $s_i^{(k)}$ at each $i$-th week as the mean of the trip scores of that week. All the values were been linearly scaled from the $[0, 100]$ range into $[0, 1]$, and hence, each $s_i^{(k)}$ takes values from 0 (bad) to 1 (excellent). The total number of weekly scores is 4419. This feature is described in the histogram in Figure 2 and in the first row of Table 2. The median score is 0.6065 and half of the values are between 0.4343 and 0.7690.
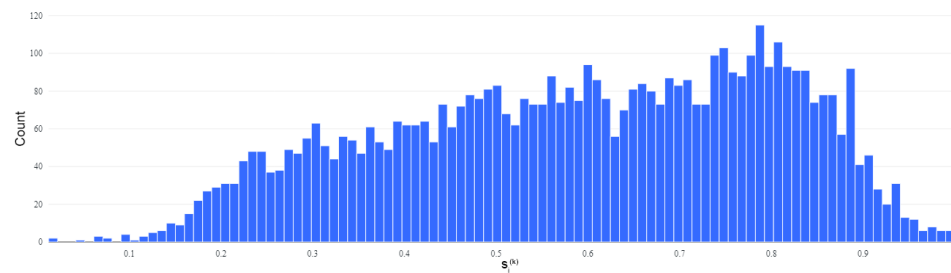
**Figure 2.** Histogram of all weekly scores.

**Table 2.** Descriptive statistics of aggregated weekly data.

| Feature | Minimum | Q1 | Q2 | Q3 | Maximum |
|---|---|---|---|---|---|
| Weekly Score | 0.0120 | 0.4343 | 0.6065 | 0.7690 | 1 |
| Weekly app session duration (s) | 0 | 0 | 12 | 186 | 4813 |
| Number of weekly app sessions | 0 | 0 | 1 | 3 | 54 |

Telematics devices, however, can also collect further behavioral features that do not (yet) go into the score. Our app also records the users' app usage by collecting data about browsing sessions. Each session is defined as a continuous time frame where the app is in the foreground on the smartphone. We used app session data to measure users' engagement. For this purpose, conceptual decisions are required. One can use the number of sessions or the time spent on the app. The number of sessions itself can be per day or per week. The time spent on the app, in turn, can be computed per session or aggregated. We opted for time spent on the app aggregated over the week.

The original dataset provided 21,283 app sessions of varying durations, up to 16 min. The app session data are presented in the second and third rows of Table 2. For each user, we first aggregated each session duration on a daily scale, then on a weekly scale, to match the processing workflow previously described for scoring. We denoted such aggregated values as $d_i^{(k)}$ for all week indexes $i = 0, 1, \dots$ and each user $k \in \{1, \dots, 498\}$. Figure 3, summarizing the behavior of our users, shows that almost all policyholders look at the app mainly in the very first weeks of the program, and they spend few minutes weekly interacting with the app.
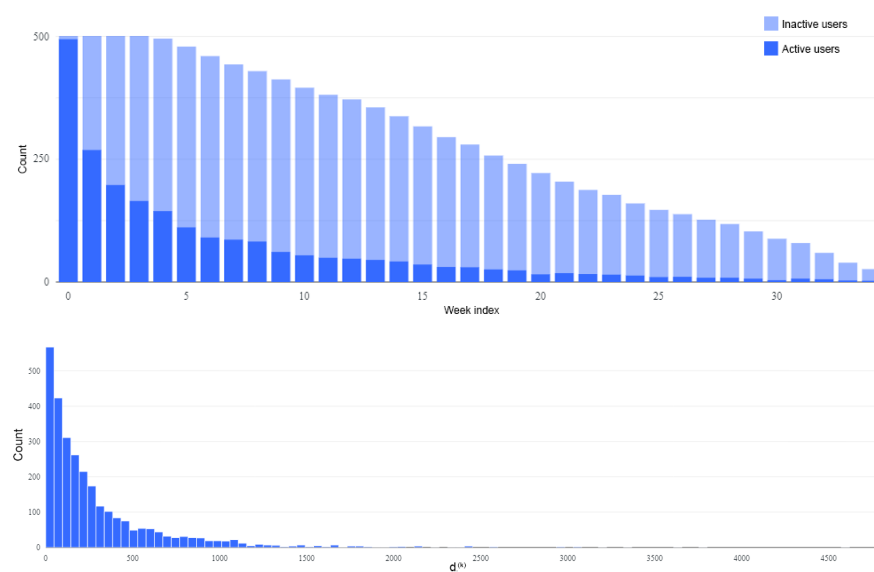


**Figure 3.** On the top, visualization of the number of active users on the telematics application (i.e., with at least one app session during the week) and inactive users (with no app sessions), as a function of the week index. On the bottom, histogram of the computed weekly durations (in seconds).

### 4.2. Approaches for Data Analysis

The first step of our analysis focused on improvement in driving behavior. In the available literature, the notion of improvement has not yet been clearly defined. Weidner, Transchel, and Weidner [31] (p. 214) claimed that neither in science nor in practice is there a "standardised method to achieve a clear 'score' of driving behavior". Soleymanian, Weinberg and Zhu in [7] aggregated scores as mean values over weeks for the entire pool of policyholders and compare them among weeks, without distinctions among users and disregarding drop-off consequences. Yet, should an increase in the average score be considered an improvement, or are there more refined ways of defining it? Should we analyze the pool with average scores, or work at individual level?

Unlike [7], we decided to analyze improvement in driving behavior for each individual policyholder, and we introduced two different workflows. The first workflow, described in Section 4.2.1, explores for each user whether there is an improvement in the initial driving style in any week of the period under consideration. The second one, discussed in Section 4.2.2, observes individual trends for the entire period.

In both cases, we split the users into four classes according to the quartile values of the 498 initial scores $s_0^{(k)}$. The *merit-based* classes, labeled as 'very-low', 'medium-low', 'medium-high', and 'very-high', represent different initial scenarios for our analysis of coaching. In order to reliably measure improvement, we reasoned, it is implausible to include all drivers in an undifferentiated group, because the margins for improvement are of course very different for bad drivers with various critical issues, which can be addressed, than for drivers who already drive excellently, for whom there is little or no room for improvement. However accurate and effective it may be, coaching will have little effect on good drivers (those we include in the medium-high and very-high groups). At the same time, we can expect it to make a difference in the driving style of very-low and medium-low groups. Our analysis of improvement, therefore, differentially explores the improvement effects in the four groups of drivers with different skill levels. A precise description of merit-based classes can be found in Table 3.

**Table 3.** Statistical characterization of the four merit-based classes, used to divide the policyholders according to their initial score $s_0$.

|  | **Very-Low** | **Medium-Low** | **Medium-High** | **Very-High** |
|---|---|---|---|---|
| Range for $s_0$ | [0, 0.4060] | (0.4060, 0.6041] | (0.6041, 0.7713] | (0.7713, 1] |
| Median score | 0.2915 | 0.4993 | 0.6853 | 0.8417 |
| Number of users | 124 | 125 | 124 | 125 |

In the following sections, we present the two approaches we used to investigate coaching effects based on two very different metrics to quantify the improvement of driving scores.

#### 4.2.1. Coaching Effects over Single Weeks

In our first set-up, we considered data points corresponding to the $s_i^{(k)}$ for i > 0. To study users' behavior after their initial week, we simply considered the difference:

$$\delta_i^{(k)} = s_i^{(k)} - s_0^{(k)}, \qquad \forall i = 1, 2, \dots \tag{1}$$

for each user *k*-th. In this case, we could independently study 3921 data points. As the scores are greater than zero, a positive value of $\delta_i^{(k)}$ denotes an enhancement in the driving style of the *k*-user in the *i*-th week with respect to his/her initial score. We consider only sufficiently high score increases given by $\delta_i^{(k)} > 0.05$ to be an *improvement*, and we associate

the corresponding data point to the *deviation-based* class 'Positive', relative to users with a positive coaching effect over single weeks. On the contrary, we cast the data points with $\delta_i^{(k)} < -0.05$ into the 'Negative' class representing weekly driving sessions with behaviors worse than the initial one. Difference values $\delta_i^{(k)} \in [-0.05, +0.05]$ correspond to a null or very moderate variation of the driving score, and the corresponding data points are therefore associated to the 'Null' class of driving sessions, with no relevant changes in the score. We note that the choice of the amplitude of the 'Null'-related range is arbitrary, and we have reasonably set it as the 5%-wide interval (as the scores are between 0 and 1, and $\delta$ can thus take values from $-1$ to $+1$).

From a methodological perspective, we remark that the definition of $\delta_i$ could also be based on the ratio $s_i^{(k)}/s_0^{(k)}$ instead of on the difference $s_i^{(k)} - s_0^{(k)}$. In that case, the three classes 'Positive', 'Negative', and 'Null' would be defined around 1 by setting as stability range the thresholds based on the 5% central interval. In our analysis, we initially took both measures into account and then opted for the difference because, based on the dataset at our disposal, this approach allowed us to better discriminate the three deviation-based classes. We associated each $\delta_i^{(k)}$ to the cumulative duration $D_i^{(k)}$ of app usage over a three-week period, which includes the current week *i* and the two preceding ones, by summing the weekly duration (in seconds) as

$$D_i^{(k)} = d_{i-2}^{(k)} + d_{i-1}^{(k)} + d_i^{(k)}, \quad \forall i \geq 2, \forall k. \tag{2}$$

The reason for including the current week of analysis and the previous two weeks is that users' behavioral patterns are not isolated within single weeks. Rather, they may be influenced by their interaction with the app in the previous weeks.

### 4.2.2. Coaching Effects over the Entire Period

The previous approach enabled us to take into account only temporary coaching effects, as the changes in each driver's score do not correlate with the passing of time. Since we are interested in long-lasting improvements as well, we also analyzed the evolution of the score with respect to the week indices *i*-th for each user independently. We implemented linear regression models:

$$s(i) := \beta_0 + \beta_1 i \tag{3}$$

explaining the score as the function *s* of *i*, for each policyholder separately. An example of linear regression is reported in Figure 4 for one (anonymous) user: on the horizontal axis, we read the index *i* of the week since the user's enrollment into the PHYD program, whereas the $[0, 1]$ vertical range is relative to the driving score function $s(i)$. His/her positive slope coefficient $\beta_1$ denotes (on average) continuously improving driving performances over all tracked weeks.
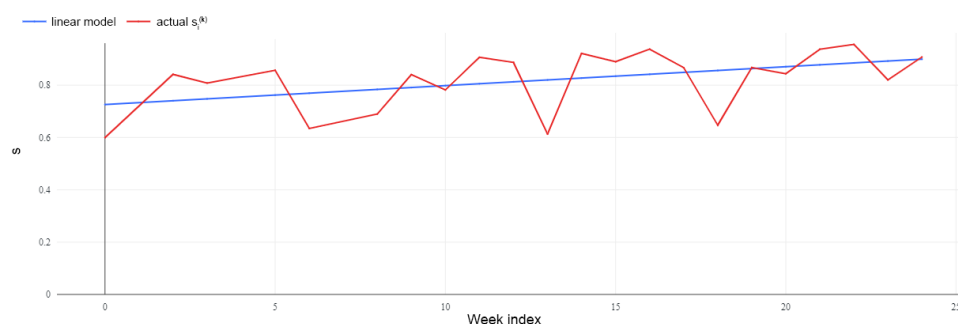


**Figure 4.** Example of simple linear regression (blue line) computed on the weekly scores (in red) as function of the week index *i*, for a user with 25 tracked weeks.

Unfortunately, many users have few weekly scores, making the regression analysis not representative for them. We thus discarded users with less than 8 scored weeks within the program and computed the regression coefficients for the remaining 212 users. We performed the statistical analyses of the $\beta_1$ coefficients, relying on the popular *p*-values, for each driver. A low *p*-value suggests that the relationship between independent and dependent variables is statistically significant, i.e., the passing of time influences the driving score within the telematic program. Conversely, a high *p*-value indicates that this relationship could plausibly be due to random fluctuations in the data rather than an actual relationship between the variables. In this regard, it is well known that the conventional significance threshold of 0.05 may be unsuitable for studies with very small sample sizes, as in our case, and there is growing acceptance in various research fields to tolerate higher thresholds, such as 0.15 or 0.20 [32].

Since our analysis focuses on capturing temporal trends rather than establishing a precise predictive model, we adopted 0.20. Thus, we interpret linear regressions with *p*-values > 0.20 as statistically not significant, and the corresponding users are categorized as 'Not Significant'. On the contrary, for the users with *p*-value $\leq$ 0.20, we can capture patterns over time by looking at the value of their slope coefficient. These drivers are divided into three classes called 'Negative', 'Null', and 'Positive', as in the previous approach. Specifically, drivers with $\beta_1 > 0.005$ are classified as 'Positive', because their (sufficiently) positive slope denotes a long-term improvement in driving behavior. Drivers with $\beta_1 < -0.005$ are classified as 'Negative', as their scores decrease over the observed weeks, while users with $-0.005 \leq \beta_1 \leq 0.005$ are classified as 'Null' to denote that no practical relevant changes have been observed in their scores over time. In the following, we refer to these clusters of users as *slope-based* classes.

The variable quantifying the engagement of each *k*-th user to the app is computed as the mean of all the weekly durations $d_i^{(k)}$, i.e., as

$$D^{(k)} = \frac{1}{N^{(k)}} \sum_{i=0}^{N^{(k)}-1} d_i^{(k)}, \tag{4}$$

where $N^{(k)}$ is the user's specific number of weeks within the telematics program. The choice to sum over all the weeks and divide by $N^{(k)}$ does not penalize policyholders who enrolled later.

### 4.3. Limitations and Remarks

Our research has some important limitations that must be taken into consideration before moving on. One limitation is that we could not assess policyholders' engagement in coaching programs that provide financial incentives. In empirical research on PHYD insurance programs, there is strong evidence that financial rewards are crucial for nudging policyholders into behavioral change [33–35]. If policyholders have no economic advantage, they apparently have little or no incentive to change their behavior. The motivation to save money is stronger than safety reason to motivate a change in behavior. There is also empirical evidence that financial incentives alone are not sufficient to motivate policyholders to engage [36]. Rewards need to be personalized, based on the risk profile of policyholders, and divided into short-term and long-term financial rewards. Unfortunately, the database we explored refers to a policy without a reward system. This lack of financial incentives can, with precision, explain the low digital engagement we observed in our database and drastically reduces the possibility of generalizing our results to the PHYD insurance market, where financial incentives are the norm.

An additional limitation of our research lies in the impossibility of linking the data on engagement and improvement of driving style with the demographic characteristics of the users, which we could not access for privacy reasons. At the beginning of our research, we wondered, for example, if there were differences in engagement between younger and older users (who presumably are less confident with digital technology), if the individual claim history was correlated with behavioral improvements and other issues. Access to this data, of course, could provide both insurance scholars and companies with important insights.

There are two further remarks, both related to time. As learned from [7], a few months may be sufficient to assess the coaching effects on the insurance pool. However, with a 9-month observation period, we could not ascertain whether and to what extent a 'habituation' effect to the insurance program might occur, i.e., whether and to what extent time could affect engagement and, consequently, the effectiveness of the coaching program. Additionally, we could not examine the behavior of policyholders over a period of time beyond the policy renewal threshold (one year and more), which could have been extremely informative but exceeds the scope of this paper.

From a methodological perspective, while we opted for a simple linear regression model to estimate individual driving score trends, we acknowledge that this choice may oversimplify behavioral dynamics. Driving style does not necessarily change linearly over time; for instance, drivers may exhibit rapid improvements during the initial weeks followed by stabilization, or they may experience cyclical fluctuations in behavior. Alternative approaches could address these aspects. Nonlinear models (e.g., polynomial regression or spline-based methods) could capture curvilinear or plateauing trends, providing a more nuanced representation of behavioral change. Similarly, time-series methods (e.g., autoregressive models) could account for temporal dependencies and periodic patterns in driving scores. However, these methods require more observations per user and introduce challenges of model standardization and interpretability across a heterogeneous user population. Given the exploratory nature of this study and the limited number of weekly observations available for each user, we adopted the linear model as a pragmatic and interpretable solution for large-scale analysis. We also note that adopting a more flexible regression model would not fundamentally alter the methodology proposed here; rather, it would extend its applicability to richer datasets. Future research could therefore incorporate nonlinear or time-series models to enhance robustness and capture more complex dynamics in driver behavior. At last, we fully acknowledge that statistical inference based on a relatively high significance threshold ($p \leq 0.20$) is unconventional and less stringent than the standard 0.05 level typically adopted in behavioral and social sciences. This decision was not made lightly: it stemmed from the exploratory nature of the study and the constraints of the dataset, where many users have only a small number of weekly observations.

We are aware of the limitations of our investigation, which depend on the type of data we had available and the constraints related to the confidentiality of our sources. Access to high-quality and high-quantity data is notoriously problematic in the field in which our investigation was conducted. However, we believe that despite these limitations, our work offers a useful contribution to the analysis of the use of behavioral data in the insurance sector. Our research highlights the importance of engagement for the analysis and implementation of coaching programs, and proposes a research methodology based on clear definitions of the concepts of engagement and coaching, which have so far been absent in the literature. The use of the available dataset enabled us to show how the proposed methodologies can be applied and how the results can be read/interpreted. The results we illustrate in session Section 5, in fact, are not meant to be representative of the

telematics motor insurance worldwide but to offer clues to analyze and understand some real dynamics of that market.

## 5. Results and Discussion

We can now outline the main results achieved based on our dataset. We first focus on improvement merely. In a second step, we analyze engagement.

### 5.1. Improvement

In our opinion, a useful definition of improvement should consider two main issues: (*i*) how to measure it and (*ii*) what variation must be detected in order to speak of improvement and, consequently, of coaching. Focusing on the first issue, we point out that even slightly different measures (i.e., difference or ratio) lead to very different interpretations of single-week improvement.

In the scatterplots of Figure 5, we look at our data points in terms of initial scores $s_0^{(k)}$ (on the horizontal axis) and $\delta_i^{(k)}$ values (on the vertical axis). In the first plot, the improvement values are computed as in Equation (1), and in the second plot, through the ratio between scores. In both plots, different colors highlight the three deviation-based classes. Especially in the first chart, one can see a wide fluctuation of scores independently from $s_0^{(k)}$, and the margins for deterioration or improvement seem to have been reached.
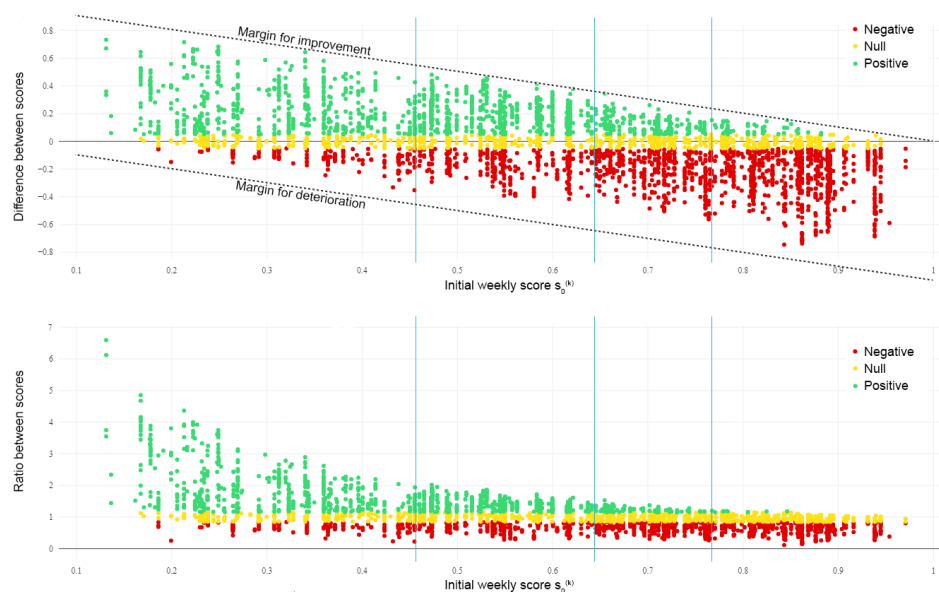


**Figure 5.** Scatterplots of data points given by $s_0^{(k)}$ initial scores, on the horizontal axis, and metrics for improvement quantification, on the vertical directions. The blue vertical lines denote the merit-based classes. On the top, the values of $\delta_i^{(k)}$ are given by difference as in Equation (1); on the bottom, the values of $\delta_i^{(k)}$ are given by ratio.

Table 4 reports the cardinality of twelve data point clusters (derived by all possible combinations among the four merit-based and the three deviation-based classes) for both approaches. The two measures of single-week deviations from the initial score led to important differences: 35.50% of data points (1390 out of 3921) denotes an improvement when it is measured as a difference, but the percentage decreases to 29.51% (1158 out of 3921) if we adopt the ratio-based approach. In any case, our data indicates an effect of improvement in the single weeks for about one third of the cases.

**Table 4.** Counting of data points in each merit-based and deviation-based class, according to two different methodologies for coaching analysis over single weeks.

| | Difference | | | Ratio | | |
|---|---|---|---|---|---|---|
| | **Negative** | **Null** | **Positive** | **Negative** | **Null** | **Positive** |
| very-low | 114 (2.91%) | 182 (4.64%) | 475 (12.10%) | 119 (3.03%) | 173 (4.41%) | 476 (12.13%) |
| medium-low | 306 (7.80%) | 215 (5.48%) | 541 (13.79%) | 265 (6.75%) | 320 (8.15%) | 477 (12.16%) |
| medium-high | 471 (12.00%) | 222 (5.66%) | 313 (7.98%) | 371 (9.45%) | 438 (11.16%) | 197 (5.02%) |
| very-high | 759 (19.34%) | 262 (6.68%) | 64 (1.63%) | 527 (13.43%) | 550 (14.02%) | 8 (0.20%) |
| Totals | 1650 (42.05%) | 881 (22.45%) | 1390 (35.50%) | 1282 (32.67%) | 1481 (37.74%) | 1158 (29.51%) |

If we instead understand improvement not as the result of single-week factors but as a positive change in driving behavior over time, we look at the slope parameter $\beta_1$ of the regression model in Equation (3) and at the slope-based classes. In Table 5, we report the cardinalities of the sixteen clusters defined according to the merit-based and the slope-based classes. First, we observe that only 39.62% of our regressions are significant. Second, most of the policyholders have significant regression slopes that are negative or very close to zero, and only 15.15% of the policyholders (30 out of 212) significantly enhanced their driving style. This means that if we observe the entire pool through individuals, on average no long-term coaching effect can be detected. However, most of the policyholders associated with the 'very-low' class have improved their driving score over weeks (12 out of 16). Finally, we note that policyholders of the 'medium-high' and 'very-high' classes still tend to be underestimated as in the previous analytic approach.

**Table 5.** Counting of data points in each merit-based and slope-based class, for coaching analysis over the entire period.

| | **Very-Low** | **Medium-Low** | **Medium-High** | **Very-High** | **Totals** |
|---|---|---|---|---|---|
| Not Significant | 28 (13.21%) | 35 (16.51%) | 30 (14.15%) | 35 (16.51%) | 128 (60.38%) |
| Negative | 3 (1.42%) | 10 (4.72%) | 15 (7.08%) | 14 (6.60%) | 42 (19.81%) |
| Null | 1 (0.47%) | 4 (1.89%) | 4 (1.891%) | 3 (1.42%) | 12 (5.66%) |
| Positive | 12 (5.66%) | 8 (3.77%) | 7 (3.30%) | 3 (1.42%) | 30 (15.15%) |
| Totals | 44 (20.75%) | 57 (26.89%) | 56 (26.42%) | 55 (25.94%) | 212 (100%) |

The division of policyholders into the four merit-based classes introduced in Section 4.2 allows us to describe our heterogeneous pool more accurately, taking into account the initial level of driving ability. We can then better interpret the real complexity and variety of the results. In Figure 6, we separately report through boxplots the $\delta_i^{(k)}$ values (computed with differences, as in Equation (1)) for the merit-based classes. For 'very-low' and 'medium-low' classes, the positive median values (represented with horizontal lines inside the orange and

red boxes) show an improvement in the driving style of policyholders. Moreover, for the 'very-low' class, the entire orange box lies over the zero line, denoting that in 75% of cases there was an increase in the score. On the other hand, as expected, the purple box is below the zero line, coherently with the numerous occurrences of red points on the right-hand side of the first scatterplot in Figure 5. Due to their limited margins for improvement, good drivers are prone to worse scores. This suggests that the 'very-high' class should not be penalized for slight score deterioration and, more generally, that any improvement or deterioration of the score should be assessed relative to the starting conditions and not in absolute terms.
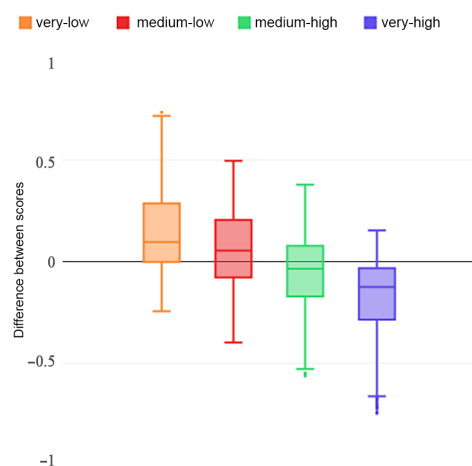


**Figure 6.** Boxplots of the $\delta_i^{(k)}$ values of Equation (1).

*5.2. Engagement*

In the last step of our analysis, we face the conceptual challenges concerning the idea of engagement. As pointed out in the Introduction, the definition of engagement is problematic. Since by 'engagement' we mean the active dealing of policyholders with the information fed back by the insurance company, we measured the duration of each user's interaction with the telematics app summing the seconds spent on the app according to the definitions given in Equations (2) and (4), which refer respectively to single weeks and the entire period. The results of the two assessments for all the considered clusters are reported in Figures 7 and 8.
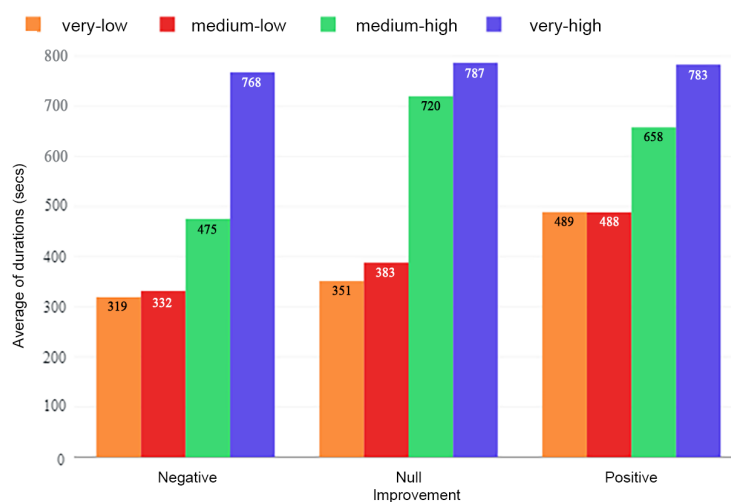


**Figure 7.** Average of the engagement metric $D_i^{(k)}$ defined in Equation (2), for the merit-based and the improvement-based classes in single weeks.
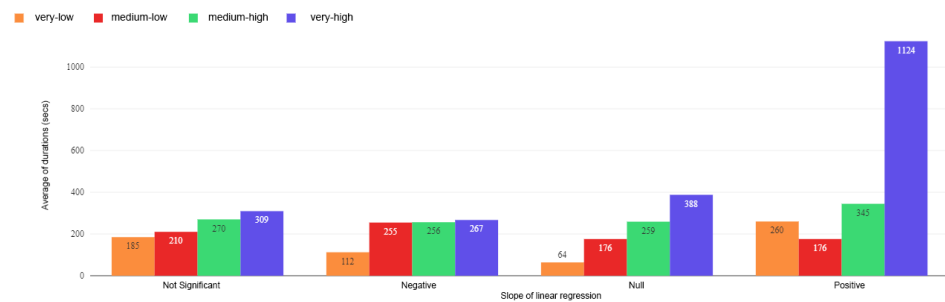
**Figure 8.** Average of the engagement metric $D^{(k)}$ defined in Equation (4), for the merit-based and the improvement-based classes over the entire period.

In both cases, it emerges that the users of the 'very-high' class (blue bars) are the most engaged with the telematics app, independently from their cluster. In most deviation- and slope-based classes, the subdivision into the four merit-based clusters reveals that the higher the initial scores, the longer the time spent on the telematics app. This holds also for the users with no statistical relevant regressions. In addition, we observe a correlation between engagement and score variation. In Figure 7, especially in the 'very-low' and 'medium-low' classes (orange and red bars), the customers manifesting positive coaching effects in single weeks are the most engaged ones, whereas less engaged customers do not improve their driving scores. In Figure 8, this trend is only partially confirmed. However, the (three) drivers in the 'very-high' class that improve their scores show very strong engagement over the entire period, well above the other groups.

Overall, these results show that there is an interplay between engagement and improvement, but this connection is not strongly evident in all cases. We suppose that the weak interplay is due to the poor engagement characterizing our dataset. Figure 3 shows that the number of users looking at the app is very low if compared to the number of users enrolled in the telematics program, and the interest in checking information feedback and scores rapidly decreases over time. Moreover, users do not have long sessions with the app, which can suggest that they do not properly read all informative pages.

## 6. Conclusions

The investigation presented in this work contributes to clarify the definition of engagement and the related methodological challenges that must be addressed when exploring engagement in proactive insurance policies. Behavioral telematics data promise to change both the traditional insurance business model and the interaction between insurers and policyholders, providing a solution to the classical problem of moral hazard [37] (p. 1231). The paradox underlying moral hazard is that policyholders are less incentivized to take precautionary measures because they are insured [38,39]. In proactive insurance, the argument goes, if individual behavior could be observed and either rewarded or penalized depending on exposure to dangers, this could affect the propensity of policyholders to control their behavior and the problem of moral hazard could be, if not removed, at least mitigated. In PHYD insurance policies, telematics of course does not control moral hazard by directly steering individual behavior. The only control that might take place is a kind of self-control, that can be based on the motivation to improve behavior, but also on the awareness of being tracked, or on the mere possibility of earning financial incentives [4]. To this purpose engagement plays a crucial role: our research shows that improvement effects are higher in policyholders who actively interact with the app.

Our findings, however, also show that engagement cannot be taken for granted. Many policyholders do not look at the app at all, and the ones who do it tend to have short and superficial sessions. The usage of any app is time-consuming and can be annoying over

time. This finding is confirmed by research in behavioral data-based health insurance, that is, in the so-called pay-as-you-live (PAYL) insurance policies [40–43]. If the goal of PHYD insurance policies is to improve driving style, policyholders must first be motivated to engage with the app in order to be motivated to change their behavior. Is there anything insurance companies can do to increase the level of engagement?

Insurance companies could expand the functionality of the app and make the interaction with it more appealing. For example, effective engagement can depend on the app usability, but also on short messages and notifications proposing challenges to achieve in order to improve driving behavior, earn points, and be rewarded. If these messages and notifications were personalized, coaching could be custom-made and the app-based interaction with the insurance company could be more exciting. Such interaction is itself a kind of behavior producing second-order behavioral data that can be recorded and used strategically. Insurance companies, therefore, could also implement second-order coaching strategies with the goal of improving engagement, besides the first-order strategies aiming at improving driving behavior.

The open question is how insurance companies can make use of these second-order scores. In principle, the level of engagement could be taken into account in the scoring process as well. Including engagement into the score, however, is not without risks. One could reward the most engaged users with points they earn when they interact with the telematics app. But users who know that their score also depends on this interaction could be inclined to game the system, learning to improve their score rather than learning to improve their driving behavior. Indeed, information fed back to the drivers can affect how people behave, but also how people deal with their behavior.

For example, a father who is late in driving his daughter to school might turn off his mobile phone and turn it back on when he drives calmly from school to work. The interaction with the app becomes strategic, if not even opportunistic. Sure, a single event does not affect a long exposure, and by means of crossing data, the lack of the trip can be detected. On the other hand, a systematic misuse can be associated to fraud. Regardless of these two extreme cases, it would be fine to reward users who stay tuned, but a reward based on the level of engagement with the app would work only if it were not disclosed. However, can insurance companies not disclose how they calculate the score?

These difficulties can be interpreted as flaws of behavioral insurance policies but also as confirmation of the crucial role of engagement, which could give rise to innovative approaches. In mundane everyday life, instead of optimizing their driving style, users are rather prone to optimize their interaction with the tracking device [40]. Instead of keeping their driving behavior under control, users keep their tracking behavior under control. This can be discouraging for software providers and insurance companies that use self-tracking technologies to implement personalized proactive prevention programs, but at the same time it is emblematic of the increasingly pervasive interaction individuals have with apps on their mobile phones. It also suggests the possibility, at least in principle, of harnessing this kind of 'addiction' in a positive manner, namely, to prevent accidents. Long-term engagement is a challenge but could also be the starting point for a more complex implementation of UBI business models by insurance companies.

**Author Contributions:** Conceptualization, E.E. and A.C.; methodology, E.M.; software, E.M. and L.R.; validation, E.E. and A.C. and E.M.; formal analysis, E.M. and L.R.; investigation, E.M., L.R. and C.M.; resources, L.R., R.T. and C.M.; data curation, L.R., R.T. and C.M.; writing—original draft preparation, E.E., A.C. and E.M.; writing—review and editing, E.E., A.C. and E.M.; visualization, E.M.; supervision, E.E.; project administration, A.C.; funding acquisition, E.E. All authors have read and agreed to the published version of the manuscript.

**Conflicts of Interest:** Authors L.R., R.T., and C.M. are data scientists and product developers at Swiss Re, which provided the proprietary database used in this study.

## References

1.　Albrecht, P. *Zur Risikotransformationstheorie der Versicherung: Grundlagen und ökonomische Konsequenzen*; Inst. für Versicherungswissenschaft, University Mannheim: Mannheim, Germany, 1990; Volume 36.

2.　Farny, D. *Versicherungsbetriebslehre*; VVW GmbH: Karlsruhe, Germany, 2011.

3.　Walters, M.A. Risk classification standards. *Proc. Casualty Actuar. Soc.* **1981**, *68*, 1–18.

4.　Cevolini, A. Coaching in Telematics Insurance: Control or Motivation? 2022. Available online: https://www.movingdots.com/news-insights/coaching-strategies-in-telematics-motor-insurance-control-or-motivation (accessed on 12 June 2024).

5.　Donaldson, G. AITE Group Report Prepared for Adept Driver Effective Driver Coaching Partnered with Telematics Improves Auto Claims and Customers Loyalty. 2020. Available online: https://www.adeptdriver.com/assets/resources/Effective-Driver-Coaching-Partnered-With-Telematics-02-21-2020.pdf (accessed on 12 June 2024).

6.　Tanninen, M.; Lehtonen, T.K.; Ruckenstein, M. Tracking lives, forging markets. *J. Cult. Econ.* **2021**, *14*, 449–463. [CrossRef]

7.　Soleymanian, M.; Weinberg, C.B.; Zhu, T. Sensor data and behavioral tracking: Does usage-based auto insurance benefit drivers? *Mark. Sci.* **2019**, *38*, 21–43. [CrossRef]

8.　Asimakopoulos, S.; Asimakopoulos, G.; Spillers, F. Motivation and user engagement in fitness tracking: Heuristics for mobile healthcare wearables. *Informatics* **2017**, *4*, 5. [CrossRef]

9.　Bucher, A. *Engaged: Designing for Behavior Change*; Rosenfeld Media: New York, NY, USA, 2020.

10.　Lupton, D. The digitally engaged patient: Self-monitoring and self-care in the digital health era. *Soc. Theory Health* **2013**, *11*, 256–270. [CrossRef]

11.　Discovery Vitality. How the Vitality Drive Programme Works. 2020. Available online: https://www.discovery.co.za/assets/discoverycoza/car-and-home-insurance/vitality-drive-terms-and-conditions.pdf (accessed on 12 June 2024).

12.　Guillen, M.; Nielsen, J.P.; Ayuso, M.; Pérez-Marín, A.M. The use of telematics devices to improve automobile insurance rates. *Risk Anal.* **2019**, *39*, 662–672. [CrossRef]

13.　Cevolini, A.; Esposito, E. From actuarial to behavioural valuation. the impact of telematics onmotor insurance. *Valuat. Stud.* **2022**, *9*, 109–139. [CrossRef]

14.　Meyers, G.; Van Hoyweghen, I. Enacting actuarial fairness in insurance: From fair discrimination to behaviour-based fairness. *Sci. Cult.* **2018**, *27*, 413–438. [CrossRef]

15.　Ericson, R.V.; Doyle, A.; Barry, D.; Ericson, D. *Insurance as Governance*; University of Toronto Press: Toronto, ON, Canada, 2003.

16.　Heimer, C.A. *Reactive Risk and Rational Action: Managing Moral Hazard in Insurance Contracts*; University of California Press: Berkeley, CA, USA, 1985; Volume 6.

17.　Stone, D. *Beyond Moral Hazard: Insurance as Moral Opportunity*; University of Chicago Press: Chicago, IL, USA, 2002.

18.　Baker, T.; Simon, J. *Embracing Risk: The Changing Culture of Insurance and Responsibility*; University of Chicago Press: Chicago, IL, USA, 2002.

19.　Litman, T. Pay-as-you-drive pricing and insurance regulatory objectives. *J. Insur. Regul.* **2005**, *23*. Available online: https://vtpi.org/jir_payd.pdf (accessed on 12 June 2024).

20.　Guillen, M.; Cevolini, A. Using risk analytics to prevent accidents before they occur—The future of insurance. *J. Financ. Transform.* **2021**, *54*, 76–83.

21.　Romanelli, L.; Albers, S. Driver coaching based on data to improve drivers' performance behind the wheel. In *Xprimm Motor Insurance Report*; XPRIMM Insurance Publications: Bucharest, Romania 2021; Volume 9, pp. 2–4.

22.　Lasry, J.-M. La rencontre choc de l'assurance et du big data. *Risques* **2015**, *103*, 19–24.

23.　Siegelman, P. Information & equilibrium in insurance markets with big data. *Conn. Ins. LJ* **2014**, *21*, 317.

24.　Chauhan, V.; Yadav, J. Bibliometric review of telematics-based automobile insurance: Mapping the landscape of research and knowledge. *Accid. Anal. Prev.* **2024**, *196*, 107428. [CrossRef]

25.　Ziakopoulos, A.; Petraki, V.; Kontaxi, A.; Yannis, G. The transformation of the insurance industry and road safety by driver safety behaviour telematics. *Case Stud. Transp. Policy* **2022**, *10*, 2271–2279. [CrossRef]

26.　Wouters, P.I.J.; Bos, J.M.J. Traffic accident reduction by monitoring driver behaviour with in-car data recorders. *Accid. Anal. Prev.* **2000**, *32*, 643–650. [CrossRef]

27.　Farmer, C.M.; Kirley, B.B.; McCartt, A.T. Effects of in-vehicle monitoring on the driving behavior of teenagers. *J. Saf. Res.* **2010**, *41*, 39–45. [CrossRef]

28. Bolderdijk, J.W.; Knockaert, J.; Steg, E.M.; Verhoef, E.T. Effects of pay-as-you-drive vehicle insurance on young drivers' speed choice: Results of a dutch field experiment. *Accid. Anal. Prev.* **2011**, *43*, 1181–1186. [CrossRef]

29. Handel, P.; Skog, I.; Wahlstrom, J.; Bonawiede, F.; Welch, R.; Ohlsson, J.; Ohlsson, M. Insurance telematics: Opportunities and challenges with the smartphone solution. *IEEE Intell. Transp. Syst. Mag.* **2014**, *6*, 57–70. [CrossRef]

30. Wahlström, J.; Skog, I.; Händel, P. Driving behavior analysis for smartphone-based insurance telematics. In Proceedings of the 2nd workshop on Workshop on Physical Analytics, Florence, Italy, 22 May 2015; pp. 19–24.

31. Weidner, W.; Transchel, F.W.G.; Weidner, R. Telematic driving profile classification in car insurance pricing. *Ann. Actuar. Sci.* **2017**, *11*, 213–236. [CrossRef]

32. Thiese, M.S.; Ronna, B.; Ott, U. P value interpretations and considerations. *J. Thorac. Dis.* **2016**, *8*, E928. [CrossRef] [PubMed]

33. Peer, S.; Muermann, A.; Sallinger, K. App-based feedback on safety to novice drivers: Learning and monetary incentives. *Transp. Res. Part F Traffic Psychol. Behav.* **2020**, *71*, 198–219. [CrossRef]

34. Stevenson, M.; Harris, A.; Mortimer, D.; Wijnands, J.S.; Tapp, A.; Peppard, F.; Buckis, S. The effects of feedback and incentive-based insurance on driving behaviours: Study approach and protocols. *Inj. Prev.* **2018**, *24*, 89–93. [CrossRef]

35. Stevenson, M.; Harris, A.; Wijnands, J.S.; Mortimer, D. The effect of telematic based feedback and financial incentives on driving behaviour: A randomised trial. *Accid. Anal. Prev.* **2021**, *159*, 106278. [CrossRef]

36. Discovery Vitality. Engagement White Paper. 2022. Available online: https://view-su2.highspot.com/viewer/62e7fa0dc9c4 8ad0468c8b65 (accessed on 12 June 2024).

37. Van Hoyweghen, I.; Horstman, K.; Schepers, R. Making the normal deviant: The introduction of predictive medicine in life insurance. *Soc. Sci. Med.* **2006**, *63*, 1225–1235. [CrossRef] [PubMed]

38. Arrow, K.J. Uncertainty and the welfare economics of medical care. In *Uncertainty in Economics*; Elsevier: Amsterdam, The Netherlands, 1978; pp. 345–375.

39. Stiglitz, J.E. Risk, incentives and insurance: The pure theory of moral hazard. *Geneva Pap. Risk Insur.-Issues Pract.* **1983**, *8*, 4–33. [CrossRef]

40. Gorm, N.; Shklovski, I. Episodic use: Practices of care in self-tracking. *New Media Soc.* **2019**, *21*, 2505–2521. [CrossRef]

41. Lupton, D. The diverse domains of quantified selves: Self-tracking modes and dataveillance. *Econ. Soc.* **2016**, *45*, 101–122. [CrossRef]

42. Ruckenstein, M.; Pantzar, M. Beyond the quantified self: Thematic exploration of a dataistic paradigm. *New Media Soc.* **2017**, *19*, 401–418. [CrossRef]

43. Tanninen, M.; Lehtonen, T.K.; Ruckenstein, M. The uncertain element: Personal data in behavioural insurance. In *Climate, Society and Elemental Insurance*; Booth, K., Lucas, C., French, S., Eds.; Routledge: Abingdon, UK, 2022; pp. 187–200.