New England University Transportation Center
77 Massachusetts Avenue, E40-279
Cambridge, MA 02139
utc.mit.edu

New England
University Transportation Center

Year 25 Final Report

Grant Number: DTRT13-G-UTC31

Project Title:

Incentivizing safer driving using peer-pressure

Project Number:	Project End Date:	Submission Date:
MITR25-12	1/31/2018	03/28/18

	Principal Investigator	Co-Principal Investigator
Investigator:	Alex “Sandy” Pentland	Erez Shmueli
Title:	Professor	Senior Lecturer, Head of Big Data Lab, Dept. of Industrial Engineering
University:	Massachusetts Institute of Technology	Tel-Aviv University, Israel
Email:	pentland@mit.edu	erez.shmueli@gmail.com
Phone:	+1-617-253-0648	+972-3-6408173

This UTC final report examines the effectiveness of monetary incentives based on real-time monitoring as means to improve driving behavior of company car drivers. We conducted a 5-months 60-drivers field study with one of the largest public transportation companies in Israel. Driving behavior was measured continuously using In-Vehicle Data Recorders (IVDR) that were pre- installed in the vehicles, enabling naturalistic, objective and concise measurements. The driving behavior measurements were then used to examine two different monetary incentive schemes: (1) a simple individual incentive scheme where each driver was rewarded based on his own improvement in driving behavior, and (2) a peer-reward scheme where each driver was rewarded based on the improvement of his peers. Drivers were also provided with daily feedback about their improvement and the reward they gained using text messages and a dedicated smartphone app. We found that the two incentive schemes presented an average improvement of 25% in driving behavior, whereas the control group (that did not use any monetary incentive) presented no improvement at all. Surprisingly, and in contrast to the reported superiority of the peer-reward scheme in previous studies, we found the individual scheme to perform better in our setting (31% vs. 15% improvement). Finally, we found that the monetary incentive schemes were able to reduce fuel consumption significantly, suggesting that such incentives can serve as a sustainable mechanism for improving driving behavior in real-world applications.

Final Publication: Y. Cohen and E. Shmueli: "Money Drives: Can Monetary Incentives based on Real-Time Monitoring Improve Driving Behavior?". Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies. Publication year: 2017

1 INTRODUCTION

Car crashes have a tremendous toll on human life and the economy. In 2013 alone, 35,200 deaths and estimated 3.8 million injuries are attributed to motor-vehicles, with an estimated cost of almost $270 billion ([32]). Human factors play a key role in the occurrence of car accidents, responsible directly for 60% of all accidents and contribute to nearly 95% of them [9].

Focusing on the case of company car drivers, the toll is known to be even more substantial. More specifically, company car drivers have an estimate chance of 20-65% to be involved in car crashes every year [12, 22, 27]. Moreover, company car drivers are 50% more likely to be involved in car crashes compared to other drivers, even after controlling for the larger distances they drive [22]. Hence, improving company car drivers' behavior may have a tremendous impact on road safety.

Measurement is considered to be the first step that leads to control and eventually to improvement [30]. With respect to driving behavior, In-Vehicle data recorders (IVDR) have recently emerged as a new and efficient tool for measuring driving behavior in a naturalistic, objective and standardized manner. Typically, an IVDR unit is a relatively small box that is installed in the vehicle and enables to monitor, record and transfer the data to a centralized computation server, in near-real-time resolution. Raw measurements of an IVDR unit typically include speed, acceleration (lateral and vertical) and GPS location ([39]). These raw measurements can then be manifested to generate series of events to characterize driving behavior.

Once the ability to measure driving behavior exists, one natural way to induce driving behavior changes is to provide drivers feedback about their measured driving behavior. A common form used to provide such feedback to company car drivers is via periodical reports. Another form that emerges recently by some IVDR manufacturers (e.g., Traffilog [35]) is to provide online feedback by means of a dashboard display or sounds. In fact, several recent studies have demonstrated that IVDR feedback can significantly improve driving behavior (see for example [39]).

Monetary incentives can serve as another means for improving driving behavior. In fact, several recent studies (see [8]) demonstrated that monetary incentives can be used to promote the driving behavior that was rewarded. Moreover, many of these studies showed that the combination of monetary incentives and feedback is more effective than feedback alone in promoting such behaviors.

In contrast to previous studies, in this paper we aim at investigating whether monetary incentives can be used to improve driving behavior of company car drivers (rather than private car owners) when using standard commercial IVDR units (instead of ad-hoc devices). In addition, while previous studies examined a single monetary incentive scheme, we compare two very different schemes. Finally, we are also interested in evaluating the impact of our monetary incentives on fuel consumption, which is a key factor in their economic sustainability.

To address these goals, we conducted a field-study with Metropoline [26], the third largest public transportation company in Israel, for a time period of 5 months, involving 60 drivers. By exploiting the pre-existing installation of IVDR units, we were able to generate driving scores for the drivers. The driving scores were then used to apply monetary incentives to encourage the drivers to improve their driving behavior. We examined two very different monetary incentive schemes: (1) a scheme based on a simple individual reward, where each driver was rewarded according to his own improvement, and (2) a peer-reward scheme that was based on inducing social pressure - each driver was rewarded according to the improvement of his peers (rather than his own improvement). Drivers were provided with daily feedback on their improvement and the rewards they gained using text messages and a smartphone app that was developed for the experiment.

Our analyses show that monetary incentives were very successful in motivating the company drivers to improve their driving behavior. While we find a significant difference between the two monetary incentive schemes used, to our surprise, the individual incentive scheme performed significantly better than the peer-reward scheme, during and after the intervention period. Finally, we show that the monetary incentives were also successful in reducing fuel consumption, suggesting that such incentives can serve as a sustainable mechanism for improving driving behavior in real-world applications.

The rest of this paper is structured as follows: In section 2 we provide the relevant background to this study and list the related work. In section 3 we give an in-depth description of the field-study that we conducted. In section 4 we present the results of our analyses and their implications in light of our research questions. In section 5 we summarize the paper and outline directions for future work.

2 BACKGROUND & RELATED WORK

Changing the behavior of road users towards safer behaviors has been a subject of much interest to researchers, practitioners and policy-makers. Traditional methods for promoting safer driving behavior relied on educational training and mass advertisement campaigns. While quite a lot of effort was invested in training and education programs, the majority of such programs was found to be lacking of formal evaluation of their effectiveness, limited in scope and power [21], and offering little in providing consistent safety benefits [4]. Additionally, there is growing skepticism among professionals and the general public about the effectiveness of mass advertisement campaigns, initiated by government and public institutes, despite the tremendous amount of money invested in them [2, 6, 25]. More recent methods rely on real-time monitoring, feedback and monetary incentives.

2.1 Real-Time Monitoring

In-Vehicle data recorders (IVDR) have recently emerged as a new and efficient tool for measuring driving behavior in an automatic, continuous, objective and standardized manner. An IVDR unit is typically a relatively small box, equipped with a rich set of sensors such as an accelerometer, GPS, a radar and a video camera. The IVDR unit is installed in the vehicle and enables to monitor, record and transfer the data to a centralized computation server, in near-real-time resolution. Raw measurements of an IVDR unit typically include speed, acceleration (lateral and vertical) and GPS location ([39]). These raw measurements can then be manifested to generate series of events that characterize driving behavior such as speeding, extreme breaking and acceleration, sharp-turns, lane departures, short headways and more [10, 11]. While IVDR systems may serve as a reliable measurement tool to study driving behavior [29, 41], their installation alone is not sufficient to significantly affect driving behavior ([16]).

2.2 Feedback

Several recent studies have shown how IVDRs can be used to provide feedback to drivers in order to improve their driving behavior. The authors of [13] and [36] showed a positive impact in driving behavior when using IVDR data to generate weekly reports that were provided to fleets' supervisors. Similarly, [41] installed IVDR systems in a fleet of small pickup trucks to track driving patterns and provided off-line reports and real-time warnings as feedback. In another study [28], the authors examined the effect of IVDR feedback on longer trips, resulting with 87% of the drivers improving their driving behavior in the monitored time period. The authors of [40] demonstrated how IVDR feedback positively affected the driving behavior of young drivers during their accompanied driving period. In another context [34], the researchers used feedback based on IVDR's data, combined with parental training in vigilant care, to reduce adolescent driving risk. The authors of [38] demonstrated lower rates of risky driving behavior by introducing parental involvement in the intervention, either by feedback or by training. [37] presented an improvement in driving skills of young drivers when using IVDR data during a marketing intervention trial in the UK. Similarly, [11] used IVDR systems to provide teenage drivers feedback on their driving behavior. A recent paper by Toledo and Shiftan [39] evaluated the impact of IVDR feedback on safe driving and eco-driving in a controlled experiment, showing that feedback can lead to a reduction of 8% in safety incidents and 3%-10% in fuel consumption.

2.3 Monetary Incentives

The possibility of promoting safer driving by means of rewards has been discussed for a long time. However, only during the last 10-15 years, technology for monitoring driving behavior has become sufficiently reliable to implement field studies that reward drivers for improving driving behavior [8]. For example, in [17, 18, 20, 33], the researchers used a punitive monetary incentive scheme based on data generated by an ISA (intelligent speeding attachment) unit, demonstrating its effectiveness in reducing speeding events. In [24], the authors examined the use of monetary incentives to reduce speeding and short headway events. They found the incentive to be quite effective, although the effect on headway maintenance was not prolonged. In [3], the authors used variable monetary incentive amounts to reduce driving mileage and found that higher amounts were more successful than lower amounts, as expected. The authors of [19] conducted a pay-as-you-speed trial that examined the combination of feedback and monetary incentive to reduce speeding. While they found each of the mechanisms alone to be useful, their combination was found to be the most effective. Commercially, we notice more and more insurance companies that reward clients for improving their driving behavior (see [42] for a list of examples).

In all of the above studies, the monetary incentive was successful in promoting the behavior that was rewarded. Moreover, many of these studies showed that the combination of monetary incentives and feedback is more effective than feedback alone in promoting such behaviors. Interestingly, all of these studies share the following aspects: (1) drivers were private car owners; (2) a dedicated ad-hoc device was used to record driving behavior, and therefore it was capable of capturing only a single type of driving events (e.g., speeding, short headway, millage and night driving); (3) subjects were recruited voluntarily, thus the researchers were likely to generate a selection bias.

In contrast, in this paper we report on a field study that was conducted with company car drivers (rather than private car owners) using standard commercial IVDR units (instead of ad-hoc devices) that were installed in the company's vehicles. As mentioned above, company car drivers are 50% more likely to be involved in car crashes compared to other drivers, and therefore improving their driving behavior may have a tremendous impact on road safety. Furthermore, an IVDR system allows us to collect data about various aspects of driving behavior. In addition, while all of the above studies examined a single monetary incentive scheme, we compare two very different schemes: an individual incentive scheme and a social incentive scheme. Finally, we also analyze the impact of our monetary incentives on fuel consumption, which is a key factor in their economic sustainability.

3 EXPERIMENTAL DESIGN

In order to test the importance of monetary incentives in improving driving behavior we conducted a dedicated field study. The field study was approved by the Institutional Review Board (IRB) and conducted under strict protocol guidelines.

When designing the field study, we had three research questions in mind: (1) can monetary incentives be used to improve driving behavior of company car drivers? (2) do different monetary incentive schemes have a different effect on the extent of improvement? and (3) can monetary incentives serve as a sustainable mechanism for improving driving behavior over time?

In the following subsections we describe the field study in detail, including how the subjects were recruited, the time periods of the study, how driving behavior score was calculated, the two monetary incentive mechanisms that were tested and the type of feedback that was provided to the subjects.

3.1 Recruiting Subjects

The subjects were recruited from Metropoline [26] - the third largest public transportation company in Israel. We limited ourselves to two bus lines operated by Metropoline (as requested by Metropoline, we omit the numbers of the two bus lines). These two lines were chosen for two reasons: (1) they have approximately the same driving routes, and therefore we do not need to control for road variability and (2) the vast majority of the buses operating in these two lines were already equipped with two different IVDR systems: ISR [5] and Traffilog [35].

We have filtered-out drivers that had less than five working days in the baseline period (the two months preceding the intervention period, see subsection 3.2), where a working day was defined as a day in which the driver drove at least two hours in the two chosen bus lines. Our filtering resulted in 63 subjects, out of which three subjects decided to drop-out at the very beginning of the experiment, resulting in a final number of 60 subjects.

,

3.2 Time Periods

The field study was divided into three time periods:

(1) **Baseline**: This period included the two months preceding the intervention period, namely June and July 2016. In this period, subjects were monitored but no intervention took place. The collected data allowed us to characterize the normal (i.e. baseline) driving behavior of subjects.
(2) **Intervention**: The actual intervention took place during the months of August and September 2016. During this period, subjects were given daily feedback about their improvement in driving behavior and the rewards they gained (according to their experimental group).
(3) **Post-intervention**: This period included the month after the intervention period, namely October 2016. During this period we continued to monitor drivers, but no feedback nor monetary incentives were used. The data that we collected in this period allowed us to better understand the after-effects of our intervention.

In order to reduce noise in our analysis, we filtered-out weekends (Friday and Saturday, the Israeli rest days) and Jewish and Muslim holidays. This filtering resulted in 42 days in the baseline period, 40 days in the intervention period and 13 days in the post-intervention period.

3.3 Experimental Conditions

The experiment included three condition groups, two groups were based on monetary incentives and one group served as control:

- **Individual Monetary Incentive**: In this group subjects were rewarded according to their own improvement in driving behavior.
- **Peer-Reward Monetary Incentive**: In contrast to the individual monetary incentive, this incentive relies on the social relationships of the subject. Subjects were shown their own progress as well as the performance of two peers that were chosen for them from the same experimental group (see Appendix A). Reward was based on the improvement of the subject's peers (rather than the subject's own improvement). We refer to the subject who is being rewarded as "mentor" and to the subjects that earn the reward for him as "trainees".
- **Control**: Subjects in this group did not receive any monetary incentive.

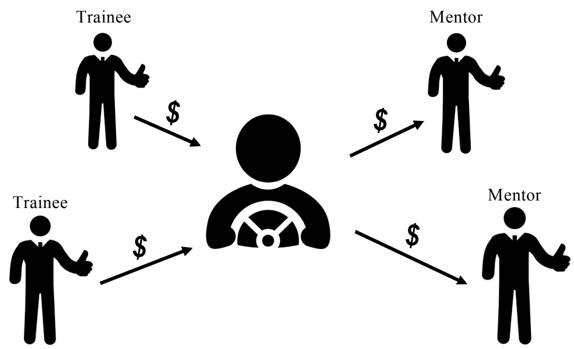


Fig. 1. An illustration of the peer-reward experimental condition. The driver in the middle serves as a mentor to the two trainees on the left (and thus is rewarded based on their improvement in driving behavior). The same driver also serves as a trainee to the two mentors on the right (hence, the two mentors on the right are rewarded based on his improvement in driving behavior).

The idea behind our selection of the two monetary incentive schemes above was to compare two very different schemes. While the individual scheme represents a simple and frequently used scheme, in which the subject's motivation to improve depends solely on the reward, the peer-reward incentive relies on social pressure induced by the mentor to increase the motivation of the trainee. A similar scheme was used in [1, 23, 31], where in all three studies, the peer-reward scheme was found to be much more effective than the individual scheme.

Out of the 63 subjects that were selected for the experiment, 23 were randomly selected to serve as the control group and the remaining 40 subjects were divided equally between the two monetary incentive groups (i.e. 20 subjects in each group). As mentioned above, at the very beginning of the intervention period, three subjects from the peer-reward group dropped-out, resulting in a total of 17 subjects in this group, 20 subjects in the individual group and 23 drivers in the control group.

3.4 The Collected Data

Our data comes from two different IVDR systems as mentioned above:

- **ISR**. Data from the ISR system was used to generate the drivers' shifts report. This report detailed for each "logical trip", the vehicle number, the driver id, and the actual start and end times of the trip. This report helped us to filter out administrative trips (i.e. going to the garage, driving back to home station) and to identify for each trip its driver.
- **Traffilog**. Data from the Traffilog system was used to generate two reports: driving alerts and fuel consumption reports. The driving alerts report detailed for each alert, the vehicle number, the event time stamp and the event category. Event categories include: moderate/hard/extreme right turn, moderate/hard/extreme left turn, moderate/hard/extreme brake and extreme acceleration). The fuel consumption report included for each "physical trip" (the period since the vehicle's engine was ignited and until the moment it was turned off) the vehicle number, the total fuel consumption in liters and start and end times of the trip.

The two reports were then joined to create a single dataset where each record was associated with a single logical trip and included the bus number, the driver id, the start and end times of the trip, the total number of alerts (and a breakdown by category) and the amount of fuel consumed.

3.5 Driving Behavior Improvement

The dataset that we described in the previous section was used to generate daily driving scores for the subjects, as well as a daily improvement score that was later translated into a monetary reward.

3.5.1 Daily Driving Behavior Score. A fundamental ingredient in our calculations is the generation of a daily driving behavior score. We chose the simplest approach according to which we just calculated the average number of alerts per driving hour that subject s had in day d, denoted by μ_s^d.

3.5.2 Baseline Driving Behavior Indicators. For each subject s, we calculated two indicators based on his daily driving behavior scores during the baseline period B:

$$\mu_s^B = \frac{\sum_{d \in B} \mu_s^d}{|B|} \tag{1}$$

$$\sigma_s^B = \sqrt{\frac{\sum_{d \in B} (\mu_s^d - \mu_s^B)^2}{|B| - 1}} \tag{2}$$

where μ_s^B is the average of all daily driving behavior scores s had during the baseline period, and σ_s^B is the corresponding standard deviation.

3.5.3 Daily Improvement Score. For each day of the intervention period, we were interested in quantifying the improvement in driving behavior of each subject in that day, compared to the baseline period. More specifically, the daily improvement score of subject s in day d of the intervention period, was calculated as:

$$\delta_s^d = (-1) * \frac{\mu_s^d - \mu_s^B}{\sigma_s^B} \tag{3}$$

Stated in words, we took the average driving behavior score of s in day d (μ_s^d), subtracted the average driving behavior score of s during the baseline period (μ_s^B), divided (normalized) it by the corresponding standard deviation (σ_s^B) and negated it (so that a positive value will represent improvement). The goal of the normalization was to reduce the effect of extremely volatile driving behavior of some of the subjects. Drivers who managed to improve their driving behavior in day d in comparison to the baseline period had a positive δ_s^d score, and those who actually got worse, had a negative δ_s^d score. The magnitude of δ_s^d depended on the extent of the improvement and the level of volatility in driving behavior of s during the baseline period.

We would like to note two points that were considered in the design of δ_s^d:

- We decided to score subjects based on their “improvement" as opposed to reducing μ_s^d beneath a predefined threshold, since: (1) we wanted both “good" drivers and “bad drivers” to be engaged in the experiment and have the option to be rewarded and (2) determining such a threshold is challenging.
- The improvement score of each subject was dependent only on himself and not on the other subjects, for two main reasons: (1) we wanted to prevent side-effects of a competition like setting in which only the top performing subjects are motivated to improve continuously whereas the others may lose hope and (2) we wanted to control for potential environmental noise, such as the sensitivity of the IVDR system, traffic variability and so on.

3.5.4 Monetary Reward. After calculating the improvement score for all subjects in a given intervention day, we determined the monetary reward for each one of them as follows. Our overall budget for the experiment allowed us to allocate 600 NIS per each day of the intervention period. This resulted, on average, in 16 NIS per day per subject (we had 37 subjects in the two monetary incentive groups). Considering the fact that the average daily salary of a driver in our setting is approximately 350 NIS, the monetary reward is equivalent to 4.5% of his salary. The monetary reward included two parts: (1) a fixed reward of 3 NIS per driver per day that was paid to all 37 subjects in all 40 days of the intervention period and (2) a variable reward that was dependent on the improvement score of the subjects in that day and their experimental condition.

The goal of the fixed reward was to encourage subjects to stay in the experiment and to continuously try to improve their driving behavior, even if they didn't manage to improve on that specific day. Moreover, even drivers who did not attend work, or did not drive the two chosen bus lines on that specific day, received the fixed reward (but were not eligible for the variable reward).

After allocating the fixed reward to the 37 drivers, the rest of the reward was divided between the two incentive groups proportionally (according to the number of active drivers in day d in each one of the two groups). We denote the daily budget in day d for each incentive group G as ω^G_d

In order to reduce frustration among the subjects, we decided to use only a positive reward and not to "punish" subjects for a "negative improvement". Therefore, all negative values of δ^d_s were truncated to 0:

$$\hat{\delta}^d_s = max(\delta^d_s, 0) \tag{4}$$

Then, in order to make the reward proportional to the improvement of the different drivers in day d, we calculated the reward per unit of improvement as follows:

$$\theta^d_G = \frac{\omega^d_G}{\sum_{s \in G} \hat{\delta}^d_s} \tag{5}$$

The total reward allocated to subject s in day d was calculated as:

$$\phi^d_s = 3 + \theta^d_G \cdot \hat{\delta}^d_s \tag{6}$$

Finally the actual reward paid to subject s was dependent on his experimental condition. In case the subject belonged to the individual group, he simply received ϕ^d_s . In case he belonged to the peer-reward group, the payment included $(\phi^d_{s_1} + \phi^d_{s_2})/2$, where s_1 and s_2 are the two trainees of s.

3.6 Additional Feedback

Besides the visual lights provided by the Traffilog IVDR system, drivers received daily feedback on their driving behavior, their improvement, and the reward they earned. Feedback was provided in two ways:

- SMS: All subjects received daily text messages to their mobile phones. In the individual group, the message included the reward the subject earned in the previous day. In the peer-reward group, the message also included a breakdown of the reward earned by each one of his trainees and the reward the subject contributed to each one of his mentors.
- A Smartphone App: we developed a dedicated smartphone application for the subjects that included detailed feedback on their driving behavior. Drivers had the option to install and use the application through a personal password that was provided to them, but were not forced to do so. The dedicated app had slightly different versions for the individual and peer-reward groups. In the 'individual' version, the home screen (Figure 2(a)) included a summary of the reward earned in the previous day, previous week, and since the beginning of the experiment. By pressing the earned amount, subjects were redirected to the trips screen (Figure 2(b)) that included information on how much money was earned at each day of the experiment, and further information about the trips made at that day, including their time, route and improvement percentages. In the 'peer-reward' version, the home screen (Figure 2(c)) showed the rewards the subject contributed to his mentors, as well as the rewards credited to him by his trainees (aggregated by the previous day, previous week, and from the beginning of the experiment). Similarly to the individual version, pressing the reward contributed to his mentors, redirected the subject to a new screen that included information

about the daily rewards he contributed to his mentors and detailed information about his trips. Pressing the reward credited by his trainees, redirected the subject to another screen (Figure 2(d)) that showed information on the daily rewards earned by the trainees for him and the trainees' trips information.

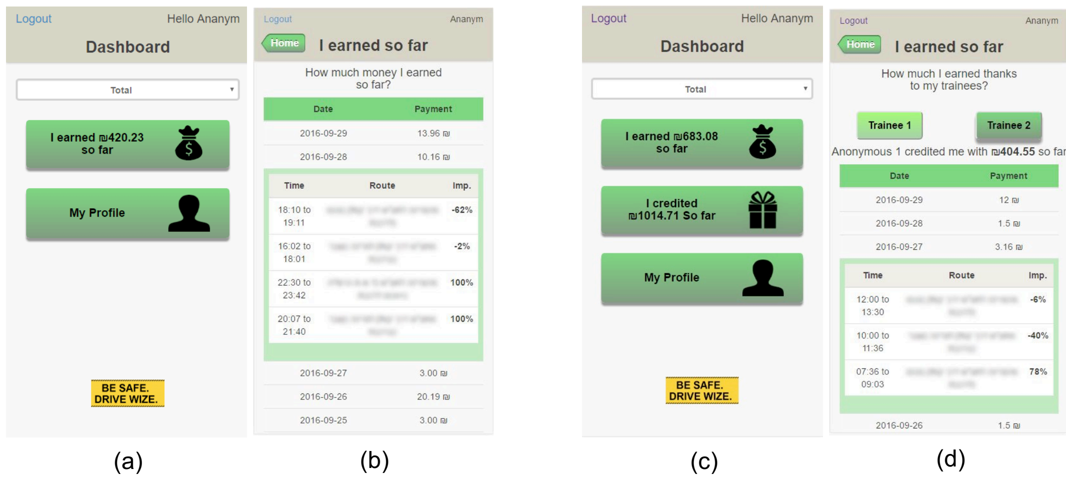


Fig. 2. Snapshots of the experimental app: (a) Individual group - Dashboard, (b) Individual group - Trips, (c) Peer-Reward group - Dashboard, (d) Peer-Reward group - Trips.

3.7 Survey

On the halfway of the intervention period we conducted a survey to collect information on how the subject perceived the experiment so far. More specifically, the survey included several parts:

- **Feedback.** This part included the following questions: "Do you receive the experiment's daily text messages to your mobile phone?", "How often do you read the experiment's text messages?", "Do you have the experiment's app installed on your mobile phone?", "How often do you use the experiment's app?".
- **Engagement and satisfaction.** This part included the following questions: "How much effort did you invest in order to improve your driving behavior?", "To what extent do you believe this experiment can change the driving patterns of Metropoline drivers?", "How satisfied are you from the reward you earn?", "What are the main reasons that made you improve your driving behavior?", "What did you do in order to improve your driving behavior?", "What are the main reasons that prevented you from improving your driving behavior further?"
- **Social interaction.** Subjects in the peer-reward group were asked additional questions: "What are the names of your two mentors?", "How close is your relationship with each one of your mentors?", "How often did your mentors encourage you to improve your driving behavior?", "How do you estimate your mentors' ability to improve your driving behavior?", "What are the names of your two trainees?", "How close is your relationship with each one of your trainees?", "How often did you encourage your trainees to improve their driving behavior?", "How do you estimate your ability to improve your trainees' driving behavior?"

3.8 Introducing the Experiment to the Drivers

The field study was described to the drivers as an experiment to improve their driving behavior. The drivers were notified that we measure their driving behavior based on Traffilog alerts (we did not provide the exact details of how driving behavior was calculated). We put a special emphasis on reminding the drivers the different categories of alerts: moderate/hard/extreme right turn, moderate/hard/extreme left turn, moderate/hard/extreme brake and extreme acceleration. Drivers were also reminded about the existence of the Traffilog light-indicator that was located above the steering wheel and provided real-time indication of alerts. We also emphasized that each day of the experiment is measured separately and independently on the previous days, and that we measure their improvement compared to themselves (during the baseline period) and independently of the other drivers. Drivers were told that they will be arbitrarily assigned to one of the two incentive groups – individual or peer-reward – and the nature of each group was explained to them. They were also notified that they cannot change their assignment once it was determined. Finally, drivers were briefed about the daily SMS messages and were encouraged to install the dedicated smartphone app on their private mobile phone.

4 RESULTS

In this section we aim at answering the three research questions that were outlined in the previous section. More specifically, in subsection 4.1 we test whether the use of monetary incentives helped in improving the driving behavior of subjects in comparison to the control group in which monetary incentives were not used. In subsection 4.2 we test whether there was a noticeable difference between the two very different monetary inventive schemes that were used. Finally, in subsection 4.3 we test the impact of the two monetary incentive schemes on fuel consumption, in order to provide some insight on the sustainability of such schemes.

4.1 The Impact of Monetary Incentives

In this subsection we compare the improvement in driving behavior between the control group and the two monetary incentive groups (that we combined into a single group for the purpose of the following analyses).

It is easy to see that our field-study fits the well-known 'pretest-posttest' experimental design. The 'pretest-posttest' design commonly includes two groups, where one group is given a treatment and the other receives no treatment ('control') over the same period of time. This design is often used in contexts such as medicine, where the effect of a new medical treatment is examined, and education, where the effect of a new teaching method is examined [7]. In our context, the treatment is the use of monetary incentives, and we examine their effect on the average driving alerts per hour (μ) during the intervention period (posttest) in comparison to the the baseline period (pretest).

The 'pretest-posttest' design aims at filtering out experimental noise and confounding variables (external validity) and allowing better examination of the intervention effect (internal validity) [7]. In particular, this design addresses a major limitation of the 'posttest only' design: it is very poor at guarding against assignment bias, because the researchers know nothing about the individual differences within the control group and how they may have affected the outcome.

Several statistical methodologies were suggested in the literature to examine the effect of the treatment in 'pretest-posttest' experimental design (each has its own pros and cons [7]). We focus here on the two popular ones: Gain Score and Mixed Design ANOVA.

Gain Score. According to this test, we compute the difference between the pretest and posttest scores for each subject and then analyze those differences in a one way ANOVA (or simply a t-test in the case of two conditional groups) using treatment (treatment vs. control) as the only factor. If the treatment main effect is significant, then the change from pretest to posttest is not the same in the two groups. However, in our setting, the difference in the average number of alerts per hour might not be suitable, as drivers with high initial alerts rate may improve more easily than drivers with low initial alerts rate. Hence, instead of calculating the difference, we calculated the improvement in percentages, denoted by Δ_s . More specifically, we first calculated ρ_s^B as the overall number of alerts divided by the overall number of driving hours, for subject s in the baseline period B. Then, we calculated

ρ_s^I as the overall number of alerts divided by the overall number of driving hours, for subject s in the intervention period I . Finally we calculate the improvement of subject s as:

$$\Delta_s = 1 \frac{\rho^I_s}{\rho^B} \tag{7}$$

Figure 3 shows the kernel density estimation (KDE) and empirical cumulative distribution function (ECDF) of the gain scores distributions of the experimental and control groups. As seen in the figure, the two distributions are near-normal and there is a significant difference in the improvement for experimental group (M=28%, SD=29%) and control group (M=-9%, SD=48%) conditions; t(58)=3.35, p-value<0.005.

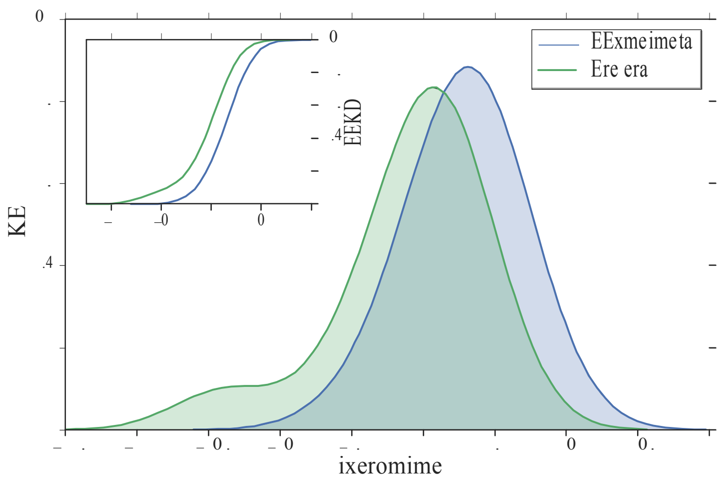


Fig. 3. KDE and ECDF of improvement for the experimental and control groups.

Mixed Design ANOVA. An alternative procedure for analyzing the pretest and posttest scores is to run a 2 x 2 ANOVA with time (pretest vs. posttest) as a within-subjects factor and treatment (treatment vs. control) as a between subjects factor. In our case, we examine the change in the average number of alerts per hour μ (score) between the baseline period and the intervention period (time) for the experimental and control groups (treatment).

The results of performing this test on our data are shown in Table 1. As shown in the table, neither time nor treatment were found to significantly affect μ, while the interaction between time and treatment was found to be significant. These results suggest that one of the groups presented a significant change in μ between the two periods, while the other did not.

Source	Type III SS	Df	Error SS	den Df	F	Pr(>F)	
(intercept)	800	1	671.79	58	69.069	1.85e-11	***
treatment	16.51	1	671.79	58	1.4251	0.237	
time	0.91	1	57.38	58	0.9237	0.340	
treatment:time	11.37	1	57.38	58	11.493	0.0012	**

Table 1. The output of the Mixed Design ANOVA test: comparing the experimental and control groups

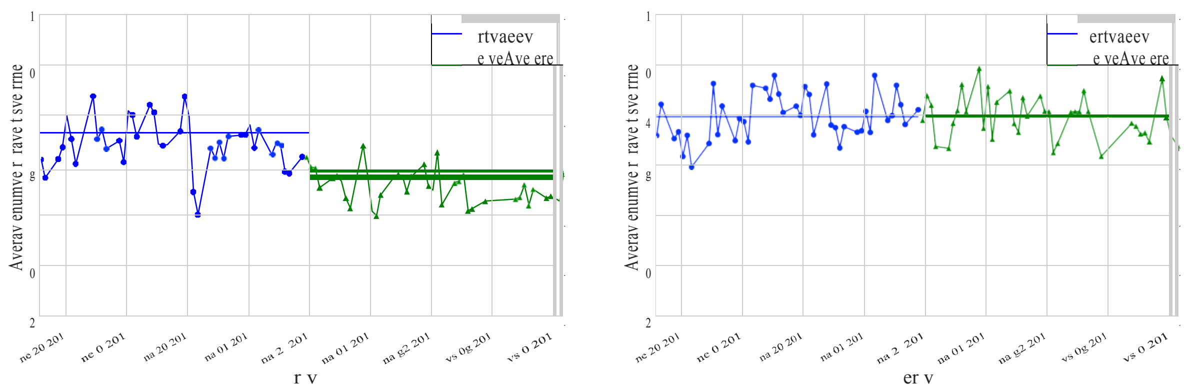
To further understand which group presented a significant change in μ, we performed additional post-hoc tests as follows. Specifically, two paired-samples t-tests were conducted to compare μ (dependent variable) in the baseline and intervention periods (independent variable) – one for the experimental group and the other for the control group. As we can see in Table 2, the experimental group showed a significant change (decrease) in μ, while the control group did not show a significant change in μ, between the baseline and intervention periods.

Experiment condition	**DF**	**t value**	**Pr> (|t|)**
Experimental	36	5.766	1.42e-06 ***
Control	22	-0.72	0.474

Table 2. The output of the two post-hoc tests - paired t-test for the difference in μ for the experimental and control groups

To further emphasize our conclusions, we calculated the overall driving score per day (for all active drivers), denoted by μ^d. μ^d is calculated by counting the overall number of alerts divided by the overall number of driving hours, for all active drivers in day *d*.

Figure 4 depicts μ^d during the baseline and intervention periods for the experimental and control groups. Each dot represents a single working day. Blue and green dots represent the baseline period and the intervention period respectively. The horizontal line in each time period indicates the mean of μ^d in that period. As can be seen from the figure, and in accordance to the previous statistical tests' findings, while in the control group the mean μ^d value barely changed between the two periods (4.05 vs. 4.02), it dropped significantly (3.66 vs. 2.79) in the experimental group. Stated differently, the experimental group presented a decrease of 22% in the average number of alerts per hour between the baseline period and the intervention period (while the control group showed no difference).



(a) Experimental group

(b) Control group

Fig. 4. Average number of alerts per hour for each day of the baseline and intervention periods.

To summarize, our findings suggest that the use of monetary incentives (regardless of the scheme) was able to improve driving behavior to a large extent.

Next, we were keen to know what was the impact of our dedicated application on the extent of improvement. More specifically, we hypothesize that drivers who used the dedicated application, tended to feel more engaged in the experiment and hence presented a more substantial driving behavioral change. In order to support this hypothesis, we examined the correlation between the number of times each subject logged-in to the application (we examined subjects in the experimental group only) and their improvement in driving behavior (i.e., Δ_s). Figure 5 shows the resulting scatter plot and regression line. We find a relatively high Pearson correlation (0.4, p-value<0.05), supporting our hypothesis that there exists a correlation between application usage and the improvement in driving behavior. Nevertheless, it is important to note this finding shows only correlation, and therefore it cannot confirm any causal relationship.

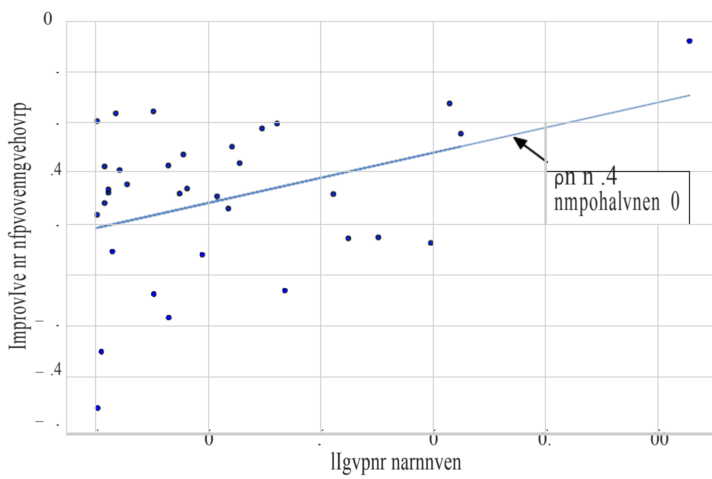
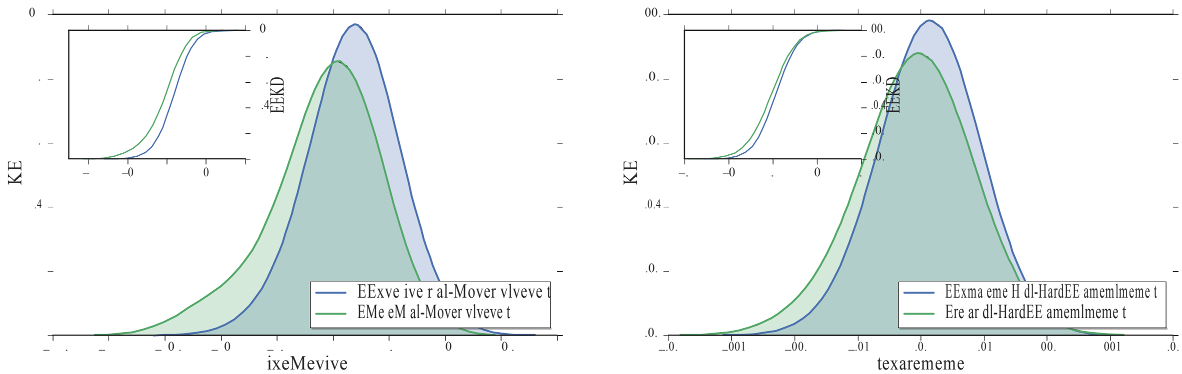


Fig. 5. Improvement as a function of application usage for subjects in the experimental group.

Recall that while Traffilog provided a severity level (moderate, hard and extreme) for each alert, our driving score (i.e., μ) treated all alerts in the same manner. Here we were interested to know whether the improvement we observed in the experimental group behaved differently for the various severity levels. In order to do so, we performed two different analyses of driving improvement (i.e., Δ_s), one for moderate severity alerts, and the other for hard and extreme severity alerts. We combined the two latter since hard and extreme alerts were quite rare. For moderate severity alerts, we find a significant difference in the improvement percentages for the experimental group (M=18%, SD=24%) and the control group (M=-8%, SD=38%) conditions; t(58)=2.9, p-value<0.01. Nevertheless, for hard-extreme severity alerts, while the mean improvement percentage for the experimental group was slightly better than that of the control group, it was not found to be significant (M=6%, SD=21% versus M=-4%, SD=31% respectively; t(58)=1.47, p-value>0.1). This finding implies that the subjects improved their driving behavior mainly by reducing their moderate-severity alerts.



(a) Moderate severity alerts

(b) Hard-Extreme severity alerts

Fig. 6. Driving behavior improvement in experimental and control groups, for (a) moderate severity alerts and (b) hard-extreme severity alerts.

4.2 A Comparison between the Two Incentive Schemes

In this subsection we report the results of a detailed comparison between the two monetary incentive schemes (i.e. individual vs. peer-reward). We follow the same methodology that was used in the previous section.

Gain score. We first calculated Δ_s for each driver in the two incentive groups. Figure 7 shows the kernel density estimation (KDE) and empirical cumulative distribution function (ECDF) of the two gain scores distributions of the two incentive groups. As seen in the figure, the two distributions are near-normal, and there is a significant difference between the mean of the improvement for individual group (M=36%, SD=25%) and peer-reward group(M=18%, SD=33%); $t(35)=1.9$, $p\text{-value}<0.1$.

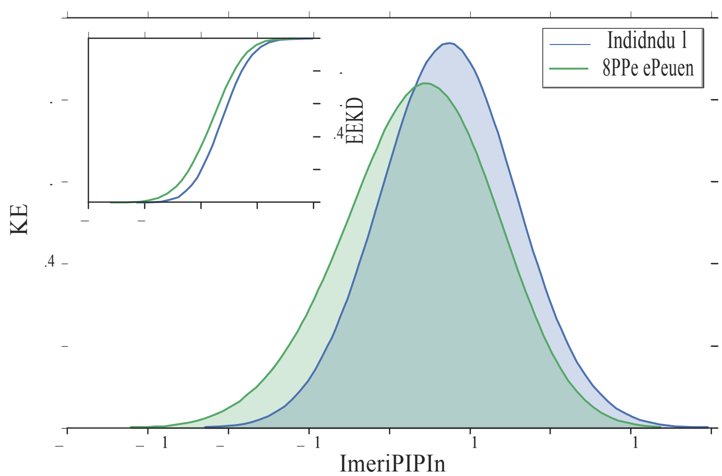


Fig. 7. KDE and ECDF of improvement for the individual and peer-reward groups.

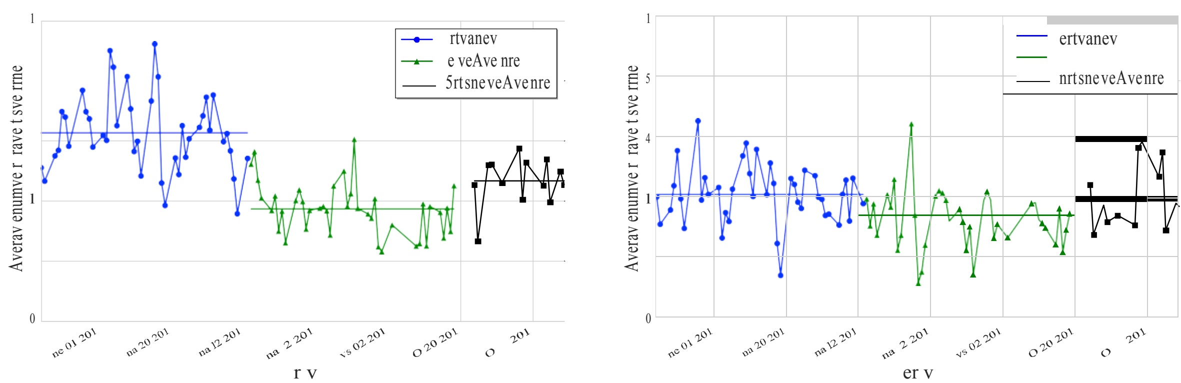
Mixed design ANOVA. We used a Mixed Design ANOVA test to examine the change in the average number of alerts per hour μ (score) between the baseline period and the intervention period (time) for the individual and peer-reward incentive groups (treatment).

The results of performing this test on our data are shown in Table 3. As shown in the table, treatment alone was not found to significantly affect μ, while time alone, and the interaction between time and treatment, were both found to be significant (p-value<0.01 and p-value<0.1 respectively). These results indicate that both groups presented a significant change in μ, and that the change presented by one of the groups was significantly higher than that of the other.

Source	Type III SS	Df	Error SS	den Df	F	Pr(>F)	
(intercept)	365.45	1	456.12	35	28.042	6.558e-06	***
treatment	1.05	1	456.12	35	0.0803	0.778	
time	3.80	1	17.84	35	7.4494	0.0098	**
treatment:time	1.57	1	17.84	35	3.0811	0.0879	.

Table 3. The output of the Mixed Design ANOVA test: comparing the individual and peer-reward incentive groups

Again, to further emphasize our conclusions, we calculated the overall driving score per day (for all active drivers) – μ^d . Figure 8 depicts μ^d in the baseline, intervention and post-intervention (marked with black squares) periods, for the individual and peer-reward groups. As can be seen from the figure, while the mean μ^d value in the individual group dropped significantly (4.12 vs. 2.86), it dropped more moderately (3.1 vs. 2.69) in the peer-reward group. Stated differently, the individual group presented an outstanding decrease of 31% in the average number of alerts per hour between the baseline period and the intervention period, while the peer-reward group showed a decrease of only 15%.



(a) Individual group

(b) Peer-reward group

Fig. 8. Average number of alerts per hour for each day of the baseline, intervention and post-intervention periods.

We also examined the effects of the two monetary incentive schemes during the post-intervention period. As can be seen in the two figures, the average alerts rate in the individual group, had risen in roughly 16%, but was still quite below that of the baseline period, meaning that subjects' continued to maintain an improved driving behavior even after stopping the monetary incentive. In contrast, the average alerts rate in the peer-reward group reached back the one presented in the baseline period. Clearly, this finding should be taken with a grain of salt due to the relatively short time of the post-intervention period.

These findings above suggest that there exists a significant difference between the two monetary incentive schemes. Surprisingly, while we expected the peer-reward scheme to outperform the individual scheme and its effect to be longer lasting (in consistency to previous studies), this was not the case here.

In order to understand better why the peer-reward incentive did not perform as well as we expected in our setting we turned to the survey that we conducted at the halfway of the intervention period. Table 4 shows five of the questions that were handed to the two incentive groups, the scale of the answers, the mean (or proportion) of the answer for each of the incentive groups. Moreover, the table reports the result of statistical tests that were conducted to check the difference in reported answers between the two groups (for the first question we used a z-test to compare proportions, and for the other four questions we used a t-test to compare means).

	Question	Individual	Peer-reward	Statistic	P-value
1	Do you have the experiment's app installed on your mobile phone? (YES/NO)	18/19	9/14	z=2.24	0.037 (**)
2	How often do you use the experiment's app? (1-5)	μ 4.39 (σ 1.19)	μ 3.89 (σ 1.45)	t(14)=0.89	0.38
3	How much effort did you invest in order to improve your driving behavior? (1-5)	μ 4.05 (σ 1.13)	μ 3.74 (σ 1.49)	t(23)=0.71	0.48
4	To what extent do you believe this experiment can change the driving patterns of Metropoline drivers? (1-5)	μ 4.52 (σ 0.84)	μ 4.28 (σ 1.2)	t(22)=0.64	0.52
5	How satisfied are you from the reward you earn? (1-5)	μ 4.37 (σ 0.95)	μ 3.38 (σ 1.04)	t(24)=2.71	0.012 (**)

Table 4. Questions that were handed to the two incentive groups as part of the survey that was conducted at the halfway of the intervention period.

Our first insight from Table 4 was that subjects in the individual group were more engaged in the experiment than subjects in the peer-reward group. More specifically, in comparison to the peer-reward group, subjects in the individual group made more effort to improve, believed harder in their potential to improve, and felt more satisfied from the reward they earned. Considering the application usage, the percentage of subjects in the individual group who used the application and their frequency of usage were higher than that of the peer-reward group. It is important to note however that only two of these five differences were found to be statistically significant (most probably due to the small sample).

Our second insight was that there was a low level of communication between mentors and trainees. In particular, in the social interaction section of the survey, we asked each subject in the peer-reward group to provide the names of his two mentors and two trainees. We found that 35% of the subjects did not remember the name of at least one of their mentors. This is especially interesting since they received daily text messages that included the name of their mentors. We found a similar result when asking the subjects for the names of their trainees - 21% of the subjects did not remember at least one of the names. We attribute these two findings to a low level of communication between the mentors and trainees. The latter can be attribute to: (1) poor social relationships between subjects to begin with and (2) low motivation to induce peer-pressure by the mentors.

Finally, we were interested to know whether the friendship level between a trainee and his mentors had an impact on his extent of improvement. Hence, we turned again to the survey, where as part of its social interaction section, we asked the subjects to rank their friendship level with their mentors (on a 1-5 Likert scale). More specifically, for each subject, we calculated the mean friendship level to his two mentors. Then, we binned the resulting friendship scores into low (beneath a score of 2.5, which was the median friendship score among all subjects) and high (a score of 2.5 and above) and compared the two corresponding distributions of improvement (see Figure 9). As can be seen in the figure, the two distributions are near-normal, and the one representing subjects with high friendship scores presents a slightly higher (but not significantly different) level of improvement. Performing a t-test, we did not find a statistically significant difference in the improvement percentages for subject with high friendship score (M=31%, SD=22%) and those with low friendship score (M=16%, SD=33%); t(35)=0.29, p-value ≈ 0.3.

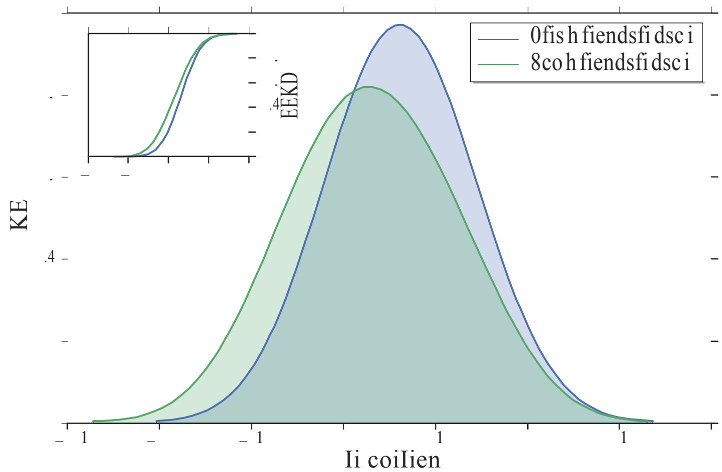


Fig. 9. KDE and ECDF of improvement for low and high friendship scores.

4.3 Sustainability

When monetary incentives are used, one question that arises is their economic sustainability. Monetary incentives require financial resources and their efficiency depends on whether their overall impact worths the monetary investment. In our case, the money that is invested in improving driving behavior can save lives, reduce car damage, lower maintenance and insurance costs, and reduce fuel consumption. In fact, several previous studies have already emphasized the benefits of safe driving as economic and fuel efficient pattern of action (e.g. [14, 15, 43]).

In the same spirit, we hypothesized that the monetary incentive schemes that we used to improve driving behavior have also led to a significant reduction in fuel consumption.

In order to test our hypothesis, we first checked whether there exists a correlation between the improvement in driving behavior of a subject (i.e., Δ_s), and his reduction in fuel consumption (calculated similarly to the improvement in driving behavior, by replacing the number of alerts with fuel consumption). Figure 10 shows the reduction in fuel consumption as a function of improvement in driving behavior, where each point is a subject in the experimental group. We find a relatively strong Pearson correlation of 0.5 (p-value<0.005).

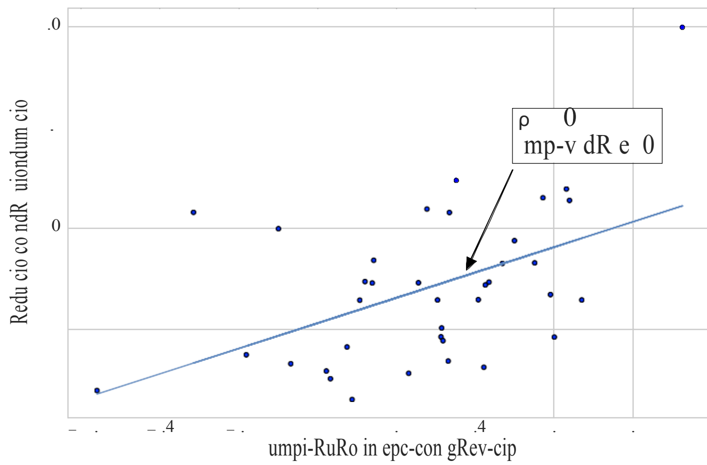


Fig. 10. Reduction in fuel consumption as a function of improvement in driving behavior.

In order to provide further support to our hypothesis, we calculated the average fuel consumption per hour (across all subjects and all days) in the baseline period as well as in the intervention period. We find that the individual incentive scheme was able to reduce 2.6% in fuel consumption per hour (from 9.13 to 8.89 liters per hour), and the peer-reward group was able to reduce 1.6% in fuel consumption per hour (from 8.84 to 8.7 liters per hour). Since, on average, a subject in the baseline period drove 5.6 hours per day and a liter of diesel costs 7.29 NIS (true to May 2017), a saving of 2.6% in liters per hour represents a saving of 9.8 NIS per subject per day[1], which is roughly 60% of the average monetary incentive per subject per day.

5 CONCLUSIONS & FUTURE WORK

This research investigated monetary incentives based on real-time monitoring as means to improve driving behavior of company car drivers. In order to do so, We conducted a 5-months 60-drivers field study in collaboration with a large public transportation company in Israel.

We found that the monetary incentive groups introduced an average improvement of 22% in driving behavior, while the control group (which did not use monetary incentives) presented no improvement at all, suggesting that monetary incentives can serve as an effective mechanism in improving driving behavior.

When we compared the two incentive schemes, we found that the individual scheme outperformed the peer-reward scheme (31% vs. 15% improvement on average). Further examining the two schemes, we found that during the post-intervention period, subjects in the individual group continued to present (a lower but significant) improvement in their driving behavior in comparison to the baseline period. In contrast to the individual group, subjects in the peer-reward group reached back their baseline driving behavior. These two findings demonstrate a clear difference between the two incentive schemes in our setting.

Inconsistent with our finding, previous studies that compared the individual and peer-reward schemes reported the peer-reward scheme to be superior to the individual scheme (e.g., [1, 23, 31]). A survey that we conducted at the halfway of the intervention period helped us to shed some light on this inconsistency, showing that subjects in the peer-reward group were less engaged in the field study and that there was a low level of communication between mentors and trainees in the peer-reward group. Moreover, it is important to note that the drivers in

[1]Detailed calculation: (9.13-8.89)*5.6*7.29=9.8

,

our field study (and most probably in other bus companies as well) have to obey a tremendously busy work schedule, and therefore don't have many opportunities to meet each other during the working day. Clearly, low engagement and low communication cannot lead to high levels of social pressure and such levels are required in order for the peer-reward scheme to become effective. Therefore, in order to leverage the peer-reward mechanism, future research should investigate ways to improve communication and friendship levels in settings like ours. Moreover, in future work we plan to examine a variant of the monetary incentive schemes used in our field-study that will combine the peer-reward and the individual incentives. Each subject will be assigned to a relatively small group, and the improvement of the subject will affect the entire group (i.e. the subject himself but also the other members of the group).

Monetary incentives require financial resources and their efficiency depends on whether their overall impact worths the monetary investment. As a first step towards demonstrating their efficiency in our case, we examined their effect on fuel consumption. We found that the individual scheme and the peer-reward scheme saved 2.6% and 1.6% respectively in fuel consumption, equivalent to roughly 60% of the monetary reward paid to the subjects. This finding is in-line with previous studies that claimed a direct relationship between safe and economic driving behavior (e.g. [14, 15, 43]). It is also important to note that improved driving behavior may have additional positive side effects on sustainability such as reduction in injuries and deaths as well as lower maintenance and insurance costs. Another way to ensure sustainability is by reducing the monetary reward. Clearly, larger amounts may better encourage improvement, but at the same time may limit the sustainability of the mechanism in real-world applications. In the future, we plan to test the effect of different amounts of monetary incentives on the overall improvement of subjects. Such investigation can help in finding the right balance.

One major limitation of this study is its scale - the length of the experiment (40 days in the intervention period and 13 days in the post-intervention period) was relatively short and the number of subjects (60) was relatively small. In the future, we plan to perform a follow-up field-study with a larger number of subjects and longer intervention and post-intervention periods. A longer intervention period will allow us to better understand the impact of seasonality on driving behavior, and a longer post-intervention period will allow us to draw clearer conclusions on the effect of monetary incentives on the longer term behavior of drivers.

Another closely related limitation is the generalizability of our findings. We examined bus drivers, operating urban routes, all working in a single company which is located in Israel. In the future, we plan to extend the study to other types of vehicles and routes, and to test whether culture plays a significant role when monetary incentives are introduced.

Finally, in this study, we limited ourselves to one single aspect of driving behavior that is related to "aggressive driving". In the future, we would like to collect data about a different aspect of driving behavior which is related to distraction. Specifically, we plan to extend our dedicated app to collect data from the drivers' smartphones regarding their usage during driving time. As noted in section 2, distractions in general, and those caused by smartphone usage in particular, are a major contributor to car crashes. The collection of such data can then be used to encourage drivers to avoid the usage of smartphones while driving and thus reduce car crashes.

ACKNOWLEDGEMENTS

The authors would like to thank prof. Alex "Sandy" Pentland (MIT Media Lab) for his help, support, advice and guidance at all stages of the experiment. We would also like to thank Isaac Hazan for his assistance with all the administrative aspects of the field study - this study could not have been conducted without his help. Finally, we would like to thank Metropoline's staff for their collaboration, always with a great smile and willingness to help.

A ASSIGNMENT OF MENTORS AND TRAINEES

The peer-reward scheme required the assignment of two mentors per each subject. In order to assign subjects to mentors in a meaningful way (i.e., in a way that is likely to induce meaningful social pressure), we performed a short survey in which we asked each subject in the peer-reward group to suggest 3-5 other subjects from the same group to serve as their mentors. Then we formulated and solved an optimization problem in which we tried to maximize the assignment of mentorships according to the suggestions of subjects.

A.1 Formulating the Assignment Problem

We denote the number of subjects in the peer-reward group as p and the the matrix of suggested mentors as A. A is of size $p \times p$, where the entry A_{ij} equals 1 if subject i suggested subject j as a mentor and 0 otherwise. Clearly, the matrix A is not necessarily symmetric.

The optimization problem tried to maximize the number of assignments that are in accordance with the mentorship suggestions of subjects, while maintaining the following constraints: (a) each subject is assigned to exactly two other mentors, (b) each subject serves as the mentors of exactly two other subjects. (c) there are no reciprocal assignment (if subject A is assigned to be the mentor of subject B, then subject B cannot be assigned as the mentor of subject A).

More formally, we searched for a matrix X of size $p \times p$ where the entry X_{ij} equals 1 if j is assigned to be the mentor of i and 0 otherwise, subjected to the following constraints:

$$
\begin{aligned}
\underset{X_{ij}}{\text{maximize}} \quad & \sum_{ij} X_{ij} * A_{ij} \\
\text{subject to} \quad & (1) \quad X_{ii} = 0, \ \forall i \\
& (2) \quad \sum_{j} X_{ij} = 2, \ \forall i \\
& (3) \quad \sum_{i} X_{ij} = 2, \ \forall j \\
& (4) \quad X_{ij} + X_{ji} \leq 1, \ \forall i, j \\
& (5) \quad X_{ij} \in \{0, 1\}
\end{aligned}
$$

where constraint (1) prevents self assignments; constraints (2) and (3) force two mentors and two trainees for each subject; (4) prevents reciprocal assignments and (5) defines the decision variable X_{ij} to be binary.

A.2 Reassignment of Mentorships

Recall that three subjects in the peer-reward group dropped the experiment at the beginning of the intervention period, causing some of the mentorship assignments of the remaining subject to be invalid. To overcome this problem, we performed a simple reassignment of mentorships as depicted in figure 11. A mentor that one of his trainees dropped the experiment, was assigned as a mentor of one of that trainee's trainees. The other mentor of the trainee that dropped out was assigned as a mentor of the second trainee of the trainee who dropped out. Clearly, this solution is not optimal in the sense that it does not take into account the mentorship suggestions of subjects. However, it is quite simple and prevents major changes in the mentorship network that may have caused great confusion among the subjects.

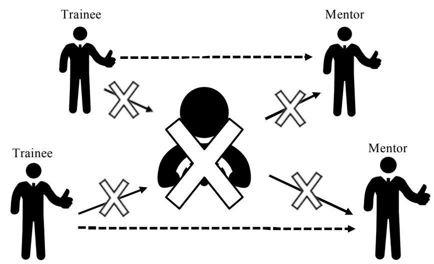


Fig. 11. Reassignment of mentorships. In case the driver in the middle of the figure dropped from the study, his two trainees were assigned to his mentor (depicted as dashed lines in the figure). Consequently, each trainee is mentor by exactly two mentors and each mentor is mentoring exactly two trainees, as required.

REFERENCES

[1] Nadav Aharony, Wei Pan, Cory Ip, Inas Khayal, and Alex Pentland. 2011. Social fMRI: Investigating and shaping social mechanisms in the real world. *Pervasive and Mobile Computing* 7, 6 (2011), 643–659.

[2] A Boulanger, M Divjak, I Orozova-Bekkevold, and V Zabukovec. 2007. Typology of Evaluation Methods: Current Practices and Campaign Evaluation. *Public Campaigns and Awareness Raising Strategies in Traffic Safety, CAST, Belgian Road Safety Institute* (2007).

[3] Jeffrey N Buxbaum. 2006. *Mileage-Based User Fee Demonstration Project: Pay-As-You-Drive Experimental Findings*. Technical Report.

[4] Ron Christie. 2001. The effectiveness of driver training as a road safety measure: An international review of the literature. In *Road Safety Research, Policing and Education Conference, 2001, Melbourne, Victoria, Australia*.

[5] ISR Integrated Systems Research Corp. 2017. (2017). http://isrfleettrack.com/ Accessed: 2017-05-04.

[6] Amanda Delaney, Bella Lough, Michelle Whelan, Max Cameron, et al. 2004. A review of mass media campaigns in road safety. *Monash University Accident Research Centre Reports* 220 (2004), 85.

[7] Dimiter M Dimitrov and Phillip D Rumrill Jr. 2003. Pretest-posttest designs and measurement of change. *Work* 20, 2 (2003), 159–165.

[8] Rune Elvik. 2014. Rewarding Safe and Environmentally Sustainable Driving: Systematic Review of Trials. *Transportation Research Record: Journal of the Transportation Research Board* 2465 (2014), 1–7.

[9] Leonard Evans. 2004. *Traffic safety*.

[10] Haneen Farah, Oren Musicant, Yaara Shimshoni, Tomer Toledo, Einat Grimberg, Haim Omer, and Tsippy Lotan. 2013. The First Year of Driving: Can an In-Vehicle Data Recorder and Parental Involvement Make It Safer? *Transportation Research Record: Journal of the Transportation Research Board* 2327 (2013), 26–33.

[11] Charles M Farmer, Bevan B Kirley, and Anne T McCartt. 2010. Effects of in-vehicle monitoring on the driving behavior of teenagers. *Journal of Safety Research* 41, 1 (2010), 39–45.

[12] H Fidderman. 1993. Company car driver behaviour. *Health and Safety Information Bulletin* 214 (1993), 9–14.

[13] Angelica C Grindle, Alyce M Dickinson, and William Boettcher. 2000. Behavioral safety research in manufacturing settings: A review of the literature. *Journal of Organizational Behavior Management* 20, 1 (2000), 29–68.

[14] Narelle Haworth and Mark Symmons. 2001. *The relationship between fuel economy and safety outcomes*. Number 188. Monash University Accident Research Centre.

[15] Patricia Hedges and Dennis Moss. 1996. Costing the effectiveness of training: case study 1-improving Parcelforce driver performance. *Industrial and Commercial Training* 28, 3 (1996), 14–18.

[16] Hans-Jürgen Heinmann and Franz-Dieter Schade. 2003. *Moderne Verkehrssicherheitstechnologie: Fahrdatenspeicher und junge Fahrer*. Wirtschaftsverl. NW.

[17] Lars Hultkrantz and Gunnar Lindberg. 2003. Intelligent economic speed adaptation. In *10th International Conference on Travel Behaviour Research, Lucerne, Switzerland*.

[18] Lars Hultkrantz and Gunnar Lindberg. 2011. Pay-as-you-speed An Economic Field Experiment. *Journal of Transport Economics and Policy (JTEP)* 45, 3 (2011), 415–436.

[19] Harry Lahrmann, Niels Agerholm, Nerius Tradisauskas, Teresa Næss, Jens Juhl, and Lisbeth Harms. 2012. Pay as You Speed, ISA with incentives for not speeding: A case of test driver recruitment. *Accident Analysis & Prevention* 48 (2012), 10–16.

[20] Gunnar Lindberg. 2006. *Valuation and pricing of traffic safety*. Ph.D. Dissertation. Örebro universitetsbibliotek.

[21] Lawrence P Lonero. 2008. Trends in driver education and training. *American journal of preventive medicine* 35, 3 (2008), S316–S323.

[22] P Lynn and CR Lockwood. 1999. The accidental liability of company car drivers. *TRL REPORT 317* (1999).
[23] Ankur Mani, Iyad Rahwan, and Alex Pentland. 2013. Inducing peer pressure to promote cooperation. *Scientific reports* 3 (2013), 1735.
[24] Undine Mazureck and Jan Hattem. 2006. Rewards for safe driving behavior: Influence on following distance and speed. *Transportation Research Record: Journal of the Transportation Research Board* 1980 (2006), 31–38.
[25] Mary Pat McKay. 2004. Unprecedented nationwide paid media campaign on impaired driving yields strong results: "You Drink & Drive. You Lose." campaign scores with target market. *Annals of emergency medicine* 44, 2 (2004), 155–156.
[26] Metropoline. 2017. (2017). https://www.metropoline.com/
[27] W Murray, D Durkin, and M Williams. 1996. *Benchmarking commercial vehicle accidents*. Technical Report. Research report, Department of Transport and Logistics, University of Huddersfield, UK.
[28] Oren Musicant, Hillel Bar-Gera, and Edna Schechtman. 2014. Temporal perspective on individual driver behavior using electronic records of undesirable events. *Accident Analysis & Prevention* 70 (2014), 55–64.
[29] Oren Musicant, Tsippy Lotan, and Tomer Toledo. 2007. Safety correlation and implications of in-vehicle data recorder on driver behavior. In *Transportation Research Board 86th Annual Meeting*.
[30] Kathryn E Newcomer. 1997. Using performance measurement to improve programs. *New directions for evaluation* 1997, 75 (1997), 5–14.
[31] Mariam Nouh, Abdullah Almaatouq, Ahmad Alabdulkareem, Vivek K Singh, Erez Shmueli, Mansour Alsaleh, Abdulrahman Alarifi, Anas Alfaris, et al. 2014. Social information leakage: Effects of awareness and peer pressure on user behavior. In *International Conference on Human Aspects of Information Security, Privacy, and Trust*. Springer, 352–360.
[32] National Safety Council. Annual Estimate of Cell Phone Crashes 2011. 2011. http://www.nsc.org/safety_road/Resources/Documents/December_2013.pdf. (2011). Accessed: 2014-04-25.
[33] Ian J Reagan, James P Bliss, Ron Van Houten, and Bryan W Hilton. 2013. The effects of external motivation and real-time automated feedback on speeding behavior in a naturalistic setting. *Human factors* 55, 1 (2013), 218–230.
[34] Yaara Shimshoni, Haneen Farah, Tsippy Lotan, Einat Grimberg, Oren Dritter, Oren Musicant, Tomer Toledo, and Haim Omer. 2015. Effects of parental vigilant care and feedback on novice driver risk. *Journal of adolescence* 38 (2015), 69–80.
[35] Traffilog Fleet Management Solutions. 2017. (2017). http://website.traffilog.com/ accessed: 2017-05-04.
[36] Beth Sulzer-Azaroff and John Austin. 2000. Does BBS work? *Professional Safety* 45, 7 (2000), 19.
[37] Alan Tapp, Ashley Pressley, Mike Baugh, and Paul White. 2013. Wheels, skills and thrills: A social marketing trial to reduce aggressive driving from young men in deprived areas. *Accident Analysis & Prevention* 58 (2013), 148–157.
[38] Orit Taubman-Ben-Ari, Sigal Kaplan, Tsippy Lotan, and Carlo Giacomo Prato. 2015. Parents' and peers' contribution to risky driving of male teen drivers. *Accident Analysis & Prevention* 78 (2015), 81–86.
[39] Galit Toledo and Yoram Shiftan. 2016. Can feedback from in-vehicle data recorders improve driver behavior and reduce fuel consumption? *Transportation Research Part A: Policy and Practice* 94 (2016), 194–204.
[40] Tomer Toledo and Tsippy Lotan. 2006. In-vehicle data recorder for evaluation of driving behavior and safety. *Transportation Research Record: Journal of the Transportation Research Board* 1953 (2006), 112–119.
[41] Tomer Toledo, Oren Musicant, and Tsippy Lotan. 2008. In-vehicle data recorders for monitoring and feedback on drivers' behavior. *Transportation Research Part C: Emerging Technologies* 16, 3 (2008), 320–331.
[42] Carmela Troncoso, George Danezis, Eleni Kosta, Josep Balasch, and Bart Preneel. 2011. Pripayd: Privacy-friendly pay-as-you-drive insurance. *IEEE Transactions on Dependable and Secure Computing* 8, 5 (2011), 742–755.
[43] Mark S Young, Stewart A Birrell, and Neville A Stanton. 2011. Safe driving in a green world: A review of driver performance benchmarks and technologies to support smart driving. *Applied ergonomics* 42, 4 (2011), 533–539.