Accident Analysis and Prevention 146 (2020) 105694

Contents lists available at ScienceDirect

Accident Analysis and Prevention

journal homepage: www.elsevier.com/locate/aap

# Traffic campaigns and overconfidence: An experimental approach☆

Thiago Christiano Silva[a,b,*], Marcela T. Laiz[a], Benjamin Miranda Tabak[c]

[a] Universidade Católica de Brasília, Distrito Federal, Brazil
[b] Department of Computing and Mathematics, Faculty of Philosophy, Sciences, and Literatures in Ribeirão Preto, Universidade de São Paulo, São Paulo, Brazil
[c] FGV/EPPG Escola de Políticas Públicas e Governo, Fundação Getúlio Vargas (School of Public Policy and Government, Getulio Vargas Foundation), Distrito Federal, Brazil

ARTICLE INFO

*Keywords:*
Traffic campaigns
Overconfidence
Video
Machine learning
Econometrics

ABSTRACT

We use a controlled experiment to analyze the impact of watching different types of educational traffic campaign videos on overconfidence of undergraduate university students in Brazil. The videos have the same underlying traffic educational content but differ in the form of exhibition. We find that videos with shocking content (Australian school) are more effective in reducing drivers' overconfidence, followed by those with punitive content (American school). We do not find empirical evidence that videos with technical content (European school) change overconfidence. Since several works point to a strong association between overconfidence and road safety, our study can support the conduit of driving safety measures by identifying efficient ways of reducing drivers' overconfidence. Finally, this paper also introduces how to use machine learning techniques to mitigate the usual subjectivity in the design of the econometric specification that is commonly faced in many researches in experimental economics.

## 1. Introduction

This paper reports on an experiment concerning the impact of watching different types of educational traffic campaign videos on overconfidence of undergraduate university students in Brazil. In our experiment, we randomly divide students into four equally-balanced groups. In the first three groups, students are assigned to watch exactly one of three different types of educational traffic campaign videos with similar length—described as technical, punitive, and shocking videos—that are intended to discourage driving under the influence of alcohol. Even though these videos have the same underlying educational message, their way of showing the consequences of driving under the influence of alcohol is distinct. After watching the video, subjects complete a questionnaire designed to assess their overconfidence, and an "overconfidence index" is constructed from the sum of their Likert-scale answers. The fourth group of students is assigned to a control group and only fill the questionnaire without watching any video. We find that students subjected to the shocking video have less overconfidence, followed by the punitive group, while the technical and control groups have roughly the same overconfidence.

Our research relies on the link between overconfidence and driving safety: by inducing changes in overconfidence through traffic video campaigns, we expect to have practical consequences to driving safety. There is a large body in the literature indicating a strong association between overconfidence and driving safety.[1] For instance, Wohleber and Matthews (2016) find that overconfidence may have essential implications for driving behavior. They show that there are significant individual differences among people that make them have different types of overconfidence and therefore driving behavior. The authors suggest the need of having different safety interventions for different types of drivers. In turn, Svenson (1981) employs an experiment and finds that drivers have a relatively high degree of overconfidence, placing themselves above the average driving peer. They believe to be more skillful and less risky than their peers. This type of overconfidence has severe consequences for driving behavior and may negatively impact the effect of traffic safety campaigns. Our paper contributes to this literature by documenting how different types of traffic campaigns videos are able to influence drivers' overconfidence and therefore impact

☆ Thiago C. Silva (Grant nos. 308171/2019-5, 408546/2018-2) and Benjamin M. Tabak (Grant nos. 310541/2018-2, 425123/2018-9) gratefully acknowledge financial support from the CNPq foundation. This research was approved by the Research Ethics Committee (*Comitê de Ética em Pesquisa*) of the *Universidade Católica de Brasília* (UCB).

* Corresponding author at: Universidade Católica de Brasília, SGAN 916 Módulo B Avenida W5, CEP 70790-160, Asa Norte, Brasília, DF, Brazil.
*E-mail addresses:* thiagochris@gmail.com (T.C. Silva), marcela.laiz@gmail.com (M.T. Laiz), benjaminm.tabak@gmail.com (B.M. Tabak).

[1] Several papers find a link between overconfidence and driving behavior. See, for instance, Rimal et al. (2019), Wohleber and Matthews (2016), Svenson (1981), and Matthews and Moran (1986).

https://doi.org/10.1016/j.aap.2020.105694
Received 11 May 2019; Received in revised form 12 March 2020; Accepted 14 July 2020
Available online 24 September 2020
0001-4575/ 

driving safety. Assessing the degree of responsiveness of drivers to different instruments that affect overconfidence is essential to think about which public policies are more effective in reducing traffic accidents.

Our paper also connects to the empirical literature that investigates how emotions and mood swings can affect the individual's decision-making process. One of the techniques—which we follow in this work—is to play a video that induces certain emotions and then evaluate how respondents who watch the video behave compared to those who do not watch the same video.[2] While we administer the questionnaire immediately after the video to ensure that we capture relative overconfidence changes caused by mood changes arising from the video content, our experiment setup does not allow to monitor respondents after the questionnaire to verify whether the overconfidence changes persist over time. While we recognize that measuring the longevity of such effect is important, we believe the effect of our experiment can be short-lived or minimal. However, our findings still have practical implications for driving safety. First, we find evidence of an effect of specific campaigns on potential driving behavior when we should expect no effect at all according to Svenson (1981). In practical terms, this result suggests that some campaigns may work better than others and therefore can support driving safety communication measures. Second, although the effect may be short-lived, the idea is that people would be exposed to these campaigns during some period.[3] We do not seek to answer whether they would have a long-lasting effect, as, in this case, we would need to expand the experiment to cover respondents over a significant period to evaluate the long-lasting effects of these campaigns. Nonetheless, we address a significant gap in the literature showing which campaigns may have a more pronounced effect on driving behavior.

To analyze the effect of different traffic campaign videos, we use a between-subject experiment as follows. We divide our sample into four *similar groups* that experience different stimuli in the form of videos. We leave one group without applying the video footage—assigned as the *control group*—such as to have a baseline behavior of overconfidence across different groups of students. In the other three groups—assigned as the *treatment groups*—we show educational traffic campaign videos with the same content but with different forms of exhibition.[4] We select three types of traffic awareness videos to encourage drivers not to drink alcohol before driving. First, a video of the Australian school, with intense and life-threatening scenes (*shocking video*). Second, a video of the American school, with punitive content, showing the consequences and penalties of drinking and driving (*punitive video*). Third, a video of the European school, with technical content, explaining the biological effects of alcohol use on the human body (*technical video*). The advantage of using videos is that they catch the attention of participants more easily and have a high degree of standardization during the laboratory procedure, allowing experiment replication (Gross and Levenson, 1995).

After watching the video content, we apply a questionnaire that aims at gauging the degree of overconfidence of students. Besides biological and sociodemographic questions, we have eight questions that assess different driving behavioral traits that can indicate excess of confidence and therefore impact driving safety.[5] We ask about the respondent's confidence in driving after drinking alcohol or while handling a cellphone. In addition, we ask respondents to self-assess their driving skill, resistance to alcohol and also alertness in handling cellphones in terms of the average driver. We then construct an *overconfidence index* based on these responses (Likert-scale answers). There is a concern about the external validity of our overconfidence estimations since our sample consists of undergraduate university students, which can have higher excess of confidence relatively to the population.[6] Our experimental design alleviates such issue by evaluating *relative changes* of the treatment to the control group. Conditionally on being similar groups, then it does not matter the actual average overconfidence level—that is argued to be higher than the population—but

[2] Similar to our experimental design, Ifcher and Zarghamee (2014) employ an experiment using mild positive and negative affects to induce mood and test how respondents change their choices after this mood-inducement experiment. They find that the mild positive affect has a positive effect on male overconfidence, whereas the effect on female participants is null. An additional important finding is that negative affects do not seem to affect overconfidence. Another paper that attempts to use video as a mood swinger is Ifcher and Zarghamee (2011), who present a film for participants and find a causal relationship between this "positive-affect" film and patience. They present a montage of Robin Williams comedy clips in 2002 (treatment) and a neutral film with landscapes and images from nature (control group). Both groups view these selected videos before making choices.

[3] Phillips et al. (2011) use a meta-analysis to test the effect of safety campaigns on roads accidents. They find a weighted average effect of −9% (95% confidence interval between 6% and 12%), with a larger effect on campaigns associated to drink-driving. The authors find that overall campaigns can be more effective in the short term when the underlying messages of the campaign are personal—i.e., they are proximal both in space and time to the behaviour that it is targeted by the campaign.

[4] Another methodological design is the within-subject design. In that, we would discard the baseline control group and administer the questionnaire twice to each subject, before and after showing the video content to the three treatment groups. In the within-subject experiment, we gain statistical power because we are able to evaluate within-individual overconfidence changes and therefore control for non-observable time-invariant characteristics of individuals (using subject fixed effects). However, within-subject designs are subject to the demand effect, in which participants may interpret the experimenter's intentions and change their behavior accordingly; thus biasing the estimates. Between-subject designs are more welcomed in experiments in which we can guarantee randomization and similarity across different groups, which are concerns we address in this work. In addition, in environments where an individual is likely to face a single decision, which is the case for driving actions, a between-subject might have more external validity. We refer the reader to Charness et al. (2012) for an interesting work that discusses these issues and compares advantages and disadvantages of between- and within-subject experimental designs. In this work, we opt for the between-subject design and mitigate potential fundamental differences between control and treatment group by controlling for several biological and sociodemographic characteristics.

[5] The measurement of confidence is clearly at the heart of our study. Merkle and Weber (2011) argue that measuring confidence is domain-specific and therefore we should mold its design considering the research question. Since we are measuring confidence in traffic, we use traffic-related questions that touch on important aspects of driving safety and self-comparison with other drivers in terms of skill, awareness, and resistance.

[6] The average student in our sample has 24 years and is single, which are traits suggesting higher levels of overconfidence comparatively to the population. For instance, Matthews and Moran (1986) perform an experiment comparing young drivers (18–25 years) with older drivers (35–50). The authors find that younger drivers are more overconfident than older drivers. They also find two relevant characteristics that are particular to this younger population. First, young drivers seem to have a dissociation between their actual skills and what they believe their skills are. Second, they tend to see themselves immune from the consequences of adopting risky decisions. Moreover, in Brazil, the insurance premium is higher for drivers up to 24 years old, suggesting that this group has more risks of getting involved in accidents than older drivers. This higher risk may be because younger drivers have not yet gone through a series of risk situations in which it was necessary to develop skills to avoid accidents. Besides, they may be more likely to use alcohol before driving or sending texts or using a cell phone during the route they are driving. The link between marital status has also been studied in the literature in the context of financial decisions. For instance, Warmath et al. (2019) show that married persons have a psychological sense of shared ownership of money and associate with lower levels of overconfidence comparatively to single persons.

the responsiveness, i.e., the changes in overconfidence due to the video content.[7]

Our sample consists of a single cross-sectional data, preventing us to insert individual fixed effects in the econometric specification to expurge non-observable time-invariant characteristics of respondents. The estimated effect of each video type relies on a differential comparison of the treatment and control groups, raising the importance of comparing similar subjects across these groups. To this end, we control for several biological and sociodemographic characteristics of respondents. To prevent discarding potentially relevant variables due a subjective and rather *ad hoc* assessment of the analyst, we use the supervised learning technique called the *elastic net*, which is a feature selection algorithm from the machine learning literature used to identify the most relevant features/attributes that best explain overconfidence.[8] There is an empirical challenge in applying such procedure in that it assumes that all the data comes from the same distribution. One underlying assumption of our experiment is that the video content has an effect on overconfidence, in such a way overconfidence distribution across different groups are expected to differ. Therefore, we first remove the *average* video effect using treatment level dummies before applying the algorithm.[9] We believe such methodology is general for dealing with multi-distributional shifts in the same sample and therefore our paper also contributes to the literature of machine learning applications in controlled experiments.

In our experiment, we use data from a large university in Brazil. The World Health Organization collects data on death on the roads of many countries around the world. Data for Brazil suggests that it ranks high among its Latin American counterparts, with more than 40,000 deaths every year (WHO, 2013, 2018). Although the country has approved strict laws restricting the use of alcohol before driving (zero tolerance), the number of deaths at the roads decreased marginally. Several campaigns try to raise awareness of the driving population, but with timid results. Understanding how to promote more effective campaigns is essential to improve these statistics and reduce deaths on the roads. Our work presents a contribution to this end.

Using the *elastic net* algorithm, we find that the top 3 features that explain overconfidence are (in order of importance): (i) whether the driver has faced a fine because of driving under the influence of alcohol, (ii) the driver's age,[10] and (iii) whether the driver has faced a fine related to handling cellphones during driving. To analyze the effect of different traffic campaign videos on overconfidence, we only keep the ten most relevant identified predictors out of the 33 analyzed. Using these predictors in an econometric exercise, we find that subjects that watched the shocking video had, on average, 8.9% less overconfidence than those in our control group. Punitive video had a relative effect of −2.1%, though not statistically significant. Technical videos had no relative effect to the control group. Since we find that our treatment and control groups are similar,[11] then we attribute these changes of overconfidence to the video content.

We also find that gender is a relevant predictor of overconfidence. In this way, we also analyze the overconfidence responsiveness of students to the different educational traffic campaign videos in terms of gender. We find that the responsiveness to punitive videos differs across genders. Punitive videos seem to be effective in reducing overconfidence for female students while they are ineffective to male students. We find that the efficacy in changing overconfidence of technical and shocking videos are unrelated to the subject's gender.

Our results provide some insights on how to design more effective public policies. A relevant finding is that how we develop safety driving campaigns matter: we find that specific videos have more effects than others. Campaigns that only provide technical explanations on how crashes in the road may occur may have little impact on overconfidence, and therefore may be innocuous to reduce accidents. We also find that demographic characteristics even matter, which suggests that we can tailor campaigns to target specific groups, such as young drivers, which are notably more prone to accidents.

## 2. Literature review

Many measures have been proposed to mitigate the number of traffic accidents and fatalities, including enhancing road capacity and design (Bucchi et al., 2012); passing stricter road safety laws, such as increasing penalties for driving while handling cellphones or under the influence of alcohol and forcing the use of seat belts, helmets and child restraints; and improving post-crash responses (WHO, 2013). However, implementation of these measures takes time and resources. This study explores alternative ways to increase road safety through the use of traffic campaign videos.[12] Our contribution lies in documenting the degree of drivers' responsiveness to different types of traffic campaign videos and how they shape drivers' attitude by inducing changes in overconfidence. Such knowledge can support the design of policies oriented to traffic safety.

A relevant question regarding the behavior of drivers and transgressions that they may commit in traffic is how they perceive the risks that they are causing to themselves and to others (Svenson, 1978). If drivers exhibit reckless behavior (overconfidence), they can commit a series of infractions making traffic more dangerous in cities, streets, and highways.[13] We can argue that a lack of confidence can also have a

[7] We should argue that in case the responsiveness to the video content differs across different population groups, then our results could be compromised in terms of external validity. For instance, in case older people respond in a more pronounced manner to shocking videos, then we would be underestimating the effect of educational traffic campaign videos with a shocking nature. This observation takes us to the importance of making within-group comparisons (same age, in the example), which is a feature that we take into account in our experimental setup, even though we do not have a considerable sample of people at advanced ages.

[8] The elastic net has the same loss function of the OLS except for the introduction of a regularization term that discourages complex models and prevents overfitting. Overfitting occurs when the algorithm learns the observed data peculiarities and not the underlying data generation distribution. While overfit model can fit well on the observed data, its external validity is poor. Elastic net contains two types of regularization to prevent such issue: the Ridge and the Lasso (Zou and Hastie, 2005). In this way, it tends to reduce the model's complexity and hence it minimizes concerns about overfitting.

[9] With this strategy, we are roughly matching the average overconfidence values of the treatment and control group by shifting the distribution of the latter to the former in average terms.

[10] We find an inverse U-shaped relation between age and overconfidence, which is consistent with the literature. For instance, Pannenberg and Friehe (2019) establish a robust relationship between age and confidence by finding empirical evidence in support of an inverted U-shape pattern to overconfidence through increasing age. They show that confidence increases up to the fifties and starts to decline thereafter.

[11] We find statistically similarity across different dimensions, such as in the share of those students that had already faced drink and cellphone fines, laterality, marital status, driving frequency, income proxy (whether she/he has a housekeeper), religiosity, age, driving experience, and mother's age.

[12] In the same vein, Chen et al. (2017) report how social comparison behavior can affect driving behavior in China. They argue that regulations combining not only traditional descriptive norms but also social status are a cost-effective yet powerful intervention for establishing driving behavior.

[13] Recent studies have also attempted to explain the behavior of drivers who handle the cellphone while driving. Existing research has attempted to explain such behavior (Tucker et al., 2015) and has identified several psychological factors associated with this risky behavior, including impulsivity (Quisenberry, 2015; Hayashi et al., 2016), habit of handling cell phones (Bayer and Campbell, 2012), cell phone dependence (Struckman-Johnson et al., 2015), and tendencies for risky behaviors (Struckman-Johnson et al., 2015). However, the behavioral and cognitive processes underlying this behavior remain unknown (Atchley et al., 2011). This trend may explain why transit legislation and traffic education have not been able to reduce the committing of this type of infraction in an effective manner (Ehsani et al., 2014; Goodwin et al., 2012).

deleterious effect on traffic safety (driving too slow on a highway may cause accidents as well).[14]

Our work connects with the empirical literature that employs experimental approaches to elucidate drivers' limitations and biases.[15] Johansson and Rumar (1968) test drivers' limitations in their visibility, whereas Cohen (1960) shows how alcoholic drink users perceive the effects of alcohol on their judgment and abilities.[16] In contrast, we seek to analyze which types of campaigns would be most effective in generating awareness of drivers in order to reduce their eventual overconfidence in traffic.

There are several papers providing evidence that it is possible to induce mood in experiment participants through the use of films.[17] Films induce changes in participants' emotional states, making them respond, on average, in different ways depending on the film content and form of exhibition. To analyze the efficacy of different traffic campaign videos, we fix the video content and explore the form of exhibition that better correlates with drivers' overconfidence changes. To the best of our knowledge, this is the first paper that performs this exercise, and we find that specific videos are more effective than others. Therefore, we show that the design of traffic campaigns that aim at reducing accidents on road matters.

There is empirical evidence that videos are capable of triggering activation in many of the emotion-related response systems (Palomba et al., 2000; Karama et al., 2002). Philippot (1993) presents a study showing how a list of 12 films reproduced six stages of different emotions, and reported success in stimulating feelings of joy, sadness, and neutrality. Gross and Levenson (1995) evaluated 16 films that provided eight mixed emotions, indicating progress in stimulating emotions: joy, anger, contentment, disgust, sadness, surprise, neutrality, and, to a lesser extent, fear. Westerman et al. (1996) analyze 11 procedures to induce mood. The literature finds that the use of movies or stories is one the most effective means of causing positive reactions in individuals. However, studies show that, while some videos are capable of inducing humor and the activation of various senses, other videos are incapable of doing that.

There is extensive use of traffic campaign videos aimed at reducing traffic accidents and alter the behavior of individuals. There is some preliminary evidence that these videos can result in reductions in traffic accidents caused by alcohol ingestion (Cameron et al., 1993; Cameron and Vulcan, 1998; Murry et al., 1993; Newstead et al., 1995), showing that non-legal sanctions help reduce this type of traffic violation. However, there is research that points out that the effects of messages encouraging people not to drink and drive may not have any impact (Agrawal and Duhachek, 2010), particularly among people less motivated to change their behavior (SEOV, 2009).[18] Our paper adds to this literature by showing that traffic campaign videos—particularly those with shocking content—are effective in reducing drivers' overconfidence.

[14] Summala (1987) studies the relationship between novice drivers and underconfidence in transit and relate traffic incidents by novice drivers to a mixture of risk taking and lack of driving skills. In turn, de Craen et al. (2008) use an adaptation test to study how novice, experienced and overconfident drivers react in different traffic scenes with increasing visual complexity. They document that novice and overconfident drivers perform worse in the adaptation test comparatively to experienced (and not overconfident) drivers. Overconfidence in this case is an incorrect self-assessment of the driver with relation to its driving skills, leading to a worse outcome in the adaptation test.

[15] For example, Shinar et al. (1980) painted stripes on a dangerous road curve, making it appear sharper than it was, causing drivers to slow down.

[16] See also Svenson (1970) and Bick and Hohenemser (1980).

[17] There is a large body of recent research that studies the impact of mood-induction, through a variety of methods, such as exposition to films, in the decision-making process (Lyubomirsky et al., 2005; Isen, 2008, 2007; Martin, 1990; Kirchsteiger et al., 2006; Rottenberg et al., 2007).

[18] See also Freeman et al. (2016).

The literature has provided evidence that most drivers believe that they are better than average. Over the years, several papers have sought to show these results empirically and in an applied way. For instance, Svenson et al. (1985) use survey with students from universities in the United States of America and Sweden and show that drivers believe that they are more skillful and safe than average and therefore tend to underestimate the risk of their behavior in traffic. This underestimation replicates a general bias of optimism that people perceive themselves to be less vulnerable than others to a variety of hazards (Weinstein, 1980). Our work also contributes to the literature by documenting that such above-the-average behavior also is the case for Brazil, an emerging market country.

## 3. Experimental design and data

In this section, we describe the experimental design and data preprocessing steps. We also provide an exploratory analysis of the experimental data.

### 3.1. Experimental design

We design a controlled experiment to investigate the effectiveness of different educational traffic campaigns videos in shaping drivers' overconfidence levels. We construct a sample that consists of students of a large university in Brasília-DF, Brazil, that hold driving licenses. We invite students across all undergraduate courses to participate in the experiment such as to obtain a representative view of population within the university. Students filled a form that consented the use of their questionnaire data for research purposes. We informed in the beginning of the experiment that participants could leave the experiment at any time. We did not provide any payments for participants and explained that they would receive a research report a few weeks after the study with overall results.

Our experiment has two sequential steps. In the first, we apply an external stimulus to respondents in the form of educational traffic campaign videos. Immediately after the video, we present a questionnaire to students querying about their characteristics, economic background, self-adherence to traffic safety measures in the case of consumption of alcoholic beverages and cell phone handling while driving, and also perceptions of their behavior while driving.

We divide our sample into four *similar groups* that experience different external stimuli in the form of videos. In the first group, no video footage is applied, and respondents only fill in the questionnaire (*control group*). In the other three groups (*treatment groups*), we show educational traffic campaign videos that discourage driving under alcohol influence but with different forms of exhibition. The objective is to induce humor change in individuals by introducing a small video before they answer the questionnaire. We select three types of traffic awareness videos to discourage driving after taking alcohol. First, a video of the Australian school, with intense and life-threatening scenes (*shocking video*). Second, a video of the American school, with punitive content, showing the consequences and penalties of drinking and driving (*punitive video*). Third, a video of the European school, with technical content, explaining the biological effects of alcohol use on the human body (*technical video*).

The advantage of using videos is that they catch the attention of participants more easily. Although threats to the standardization are present in any laboratory procedure, the content of the stimulus, presentation apparatus and viewing conditions are under tight control with the application of videos. The standardization of videos is therefore high, allowing the potential replication of effects between laboratories (Gross and Levenson, 1995). High uniformity, however, does not guarantee that the mood-changing impact of the video will be the same for all participants.

All steps in the experiment are identical, except for the video content to which students are submitted. Videos are about the same length so that video length is not the mood-inducing factor, but the content and presentation of the video. At the beginning of each experiment, students are instructed not to perform any communication with each other, so that results are not biased. The questionnaire is answered by participants immediately after the video, in such a way that no time lapse would hinder the capture of humor provoked by the video.

To mitigate fundamental differences of the control and treatment groups, which could otherwise compromise our results, we randomly assign students into the treatment and control groups within classes from the same undergraduate course *and* period.[19] We expect courses from the same undergraduate course and the same period be roughly homogeneous in the aggregate. In this way, control and treatment groups should be approximately equivalent *ex-ante* the application of the external stimulus (video) and selection bias should be minimal.

As robustness, we also construct similarity clusters and compare individuals within these clusters that are at various courses in the same period, in which one was assigned to the control group and the other to the treatment group. We use social and bio-characteristics to compose these groups such as age similarity, laterality, marital status, whether it is a frequent driver and suffered drink and cellphone-related fines, age differential between the mother and the student, and self-declaring emotion level during the questionnaire.

1

## 3.2. Data

The sample size is an essential concern in behavioral economics experiments, particularly for studies involving controlled trials (List et al., 2011). We choose the sample size using *a priori analysis*. Such analysis provides an efficient method for controlling the statistical power that one conducts a survey (Cohen, 1988). We use the well-established and freely available G*Power tool to determine the required sample size (Faul et al., 2007). In our case, we base our sample size choice on the usual benchmark in behavioral research: a significance level of 0.05, power adjustment of 0.80, and size effect of 0.5 (List et al., 2011). Under these criteria, we obtain a minimum sample size of about 84 participants.

The experiment is applied to a sample of 400 students, equally divided into four groups: 100 questionnaires administered to the control group and 100 surveys applied to each of the three treatment groups. Respondents are between 19 and 58 years old, 183 (45.8%) are women, and 217 (54.2%) are men. We implemented the experiment in 2018. Respondents had 15 min to respond to the questionnaire after the application of the video.

In the following, we discuss the grave concern about the similarity between the control and treatment groups and also define our overconfidence index.

### 3.2.1. Similarity between the control and treatment groups

One potential concern in our analysis is selection bias, which is the existence of fundamental differences between the control and treatment groups. Selection bias hampers any causal analysis and is a live concern in any empirical study. A way to minimize the selection bias across these groups is through randomization. In this section, we discuss the composition of the control and treatment groups and show that they have statistical similarity. This fact suggests that our randomization procedure has worked as planned.

Table 2 compares social, biological, economic and driving characteristics across the control and the three treatment groups that experienced video exhibitions. We show that these groups are almost identical in these characteristics, suggesting that our randomization strategy has worked. We also test for statistical differences across the multiple groups for each attribute and report the $p$-value in the last column. The null hypothesis is that the distributions are statistically the same. As we can see, we cannot reject the null hypothesis in all cases except for the gender. Therefore, in our robustness, we will take care of this heterogeneity by also comparing within gender (within estimation).

The proper homogenization between the control and treatment groups allows us to test the effect of different traffic campaign videos in shaping drivers' overconfidence.

### 3.2.2. Construction of the overconfidence index

We construct our overconfidence based on eight questions that are highlighted in the first column of Table 3. We score each question based on the five possible and standardized answers: strongly agree (+2 points), agree (+1 point), indifferent (no points), disagree (−1), and strongly disagree (−2 points). The overconfidence index is the sum of points achieved by respondents in the eight questions. Respondents with positive overall scores have a degree of overconfidence, while respondents with a negative overall score are underconfident. Zero scores denote average individuals in terms of self-confidence.

Table 3 provides the average scores for each question (row) and group (column), which can range from −2 (strong underconfidence) to +2 (strong overconfidence). Since most scores are positive, respondents have traits of overconfidence, on average, regardless of their assigned groups. In addition, the overall overconfidence (last row) is particularly higher for the control group and the technical video treatment group, suggesting that traffic educational videos with technical content might not be efficient to reduce drivers' overconfidence. The level of overconfidence significantly drops for respondents assigned to the shocking video treatment group, hinting us that educational videos with shocking content may reduce drivers' overconfidence. Punitive videos are in-between technical and shocking videos in terms of reducing drivers' overconfidence. We will test these conjectures using an econometric specification later, in which we control for several confounding factors that could be driving these plain averages.

Fig. 1 depicts boxplots of the overconfidence indices for different social and bio-characteristics of respondents such as (a) gender, (b) marital status, (c) laterality, and (d) driving experience. With no external intervention (control group), we see that women are more overconfident than men in traffic (statistically significant at 5% level), right-handed persons are more overconfident than left-handed persons (statistically significant at 1% level), and drivers with medium or high experience are more overconfident than drivers with low experience (statistically significant at 1% level). We do not find significant differences between single and married persons in the control group. Consistent with Table 3, the median overconfidence indices tend to decrease when we look at the treatment groups that watched the technical, punitive, and shocking videos (in this order).

## 4. Feature selection using machine learning

Our questionnaires collect 33 social, biological and economic characteristics of respondents. Many of these attributes may be correlated with each other, which could inflate our standard errors in our econometric analysis exploring the determinants of drivers' overconfidence. To avoid such problem and also to not miss out any important covariate able to strongly explain drivers' overconfidence, we use data-driven machine learning methods to help us select the most important attributes. We only keep attributes identified as most important by the machine learning method that have economic meaning, so as to prevent spurious correlations.

By assumption, we assume that students from the control and treatment groups are roughly similar. However, they are exposed to different external stimuli in the form of educational traffic videos. In this way, our outcome variable—the level of overconfidence—is

[19] Our external intervention in the form of videos is at the class level, i.e., all students from the same class watch the same video. Therefore, we cannot assign students among different groups within the same class. In this way, we assign them across different classes in the same period within the same university, such as to alleviate concerns with selection biases.

**Table 1**
We compare social, biological, and driving characteristics across the control and the three treatment groups. For categorical variables, we report the number of occurrences followed by the associated percentage in parentheses. For numerical variables, we provide the mean value followed by the associated standard deviation. For each variable, we also test whether the control and treatment group statistics are statistically equivalent and report the associated p-value in the last column. In this comparison, we perform pairwise comparisons adjusting for multiple testing (Tukey when row-variable is normal-distributed and Benjamini & Hochberg method otherwise). The $p$-value is computed from the Pearson test when row-variable is normal and from the Spearman test when it is continuous non-normal.

| | No video $N = 100$ | Technical video $N = 100$ | Punitive video $N = 100$ | Shocking video $N = 100$ | Overall $p$-value |
|---|---|---|---|---|---|
| *Categorical variables* | | | | | |
| Drink fines? | | | | | 0.851 |
| No | 97 (97.0%) | 94 (94.0%) | 96 (96.0%) | 95 (95.0%) | |
| Yes | 3 (3.00%) | 6 (6.00%) | 4 (4.00%) | 5 (5.00%) | |
| Cellphone fines? | | | | | 0.613 |
| No | 89 (89.0%) | 91 (91.0%) | 88 (88.0%) | 85 (85.0%) | |
| Yes | 11 (11.0%) | 9 (9.00%) | 12 (12.0%) | 15 (15.0%) | |
| Laterality | | | | | 0.546 |
| Left-handed | 7 (7.00%) | 9 (9.00%) | 10 (10.0%) | 13 (13.0%) | |
| Right-handed | 93 (93.0%) | 91 (91.0%) | 90 (90.0%) | 87 (87.0%) | |
| Marital status | | | | | 0.719 |
| Married | 16 (16.0%) | 17 (17.0%) | 20 (20.0%) | 14 (14.0%) | |
| Single | 84 (84.0%) | 83 (83.0%) | 80 (80.0%) | 86 (86.0%) | |
| Frequenty driver? | | | | | 0.249 |
| No | 34 (34.0%) | 29 (29.0%) | 28 (28.0%) | 40 (40.0%) | |
| Yes | 66 (66.0%) | 71 (71.0%) | 72 (72.0%) | 60 (60.0%) | |
| Gender | | | | | <0.001 |
| Female | 59 (59.0%) | 34 (34.0%) | 32 (32.0%) | 58 (58.0%) | |
| Male | 41 (41.0%) | 66 (66.0%) | 68 (68.0%) | 42 (42.0%) | |
| Has housekeeper? | | | | | 0.131 |
| No | 85 (85.0%) | 93 (93.0%) | 82 (82.0%) | 85 (85.0%) | |
| Yes | 15 (15.0%) | 7 (7.00%) | 18 (18.0%) | 15 (15.0%) | |
| Is religious? | | | | | 0.624 |
| No | 37 (37.0%) | 31 (31.0%) | 39 (39.0%) | 33 (33.0%) | |
| Yes | 63 (63.0%) | 69 (69.0%) | 61 (61.0%) | 67 (67.0%) | |
| *Numerical variables* | | | | | |
| Age | 25.7 (5.97) | 25.7 (5.82) | 27.0 (6.48) | 26.8 (7.37) | 0.313 |
| Driving experience | 5.29 (6.87) | 5.66 (4.40) | 5.31 (5.21) | 4.25 (3.50) | 0.242 |
| Mother's–student's age | 24.7 (6.70) | 25.0 (5.24) | 24.3 (5.44) | 23.9 (4.32) | 0.518 |
| Emotion | 4.45 (1.57) | 4.55 (1.56) | 4.56 (1.39) | 4.79 (1.34) | 0.413 |

**Table 2**
We use eight components to gauge drivers' overconfidence. For each element, the respondent can answer strongly agree with the assertion (+2 points for overconfidence), agree (+1), indifferent (0), disagree (−1), and strongly disagree (−2). We calculate the overall as the average of each of these components, and we report it in the last row of the table. Each intermediate cell represents the average for the specific question (row) and group column) and can range from −2 (underconfidence) to +2 (overconfidence).

| Assertion (question) | No video | Technical video | Punitive video | Shocking video |
|---|---|---|---|---|
| I am confident in driving after drinking alcohol | 0.19 | 0.08 | 0.27 | −0.01 |
| I am confident in driving while in cellphone | 0.21 | 0.12 | 0.15 | 0.08 |
| I feel like a driver above the average | 0.76 | 0.65 | 0.48 | 0.30 |
| I can drive after ingesting alcohol without major risks because my resistance is above the average | 0.05 | 0.11 | 0.06 | −0.02 |
| I feel that I am more careful and respectful of traffic laws than the average driver | 0.81 | 0.86 | 0.75 | 0.50 |
| I feel that I am more aware of traffic than the average driver | 0.83 | 0.95 | 0.79 | 0.50 |
| I do not see any problem driving under the influence of alcohol because I am quite resistant | 0.02 | 0.08 | 0.02 | 0.03 |
| I do not see any problem driving while in cellphone because I am very alert in traffic | 0.03 | 0.08 | 0 | 0.03 |
| **Overall** | 0.181 | 0.183 | 0.158 | 0.088 |

somewhat contaminated by such exposure. In any machine learning method, data must come from the same distribution and any shifting of the distribution is not appropriate when training and can lead to biased results. As a first step, we remove the effect of such external stimuli used in the design of the control and treatment group by running the following regression:

$$y_i = \alpha + \beta_1 \,\mathrm{Technical}_i + \beta_2 \,\mathrm{Punitive}_i + \beta_3 \,\mathrm{Shocking}_i + \epsilon_i, \tag{1}$$

in which $y_i$ denotes the level of overconfidence of student $i$, $\mathrm{Technical}_i$, $\mathrm{Punitive}_i$, and $\mathrm{Shocking}_i$ are non-overlapping dummy variables that are switched on if student $i$ received the corresponding treatment (technical, punitive, or shocking video), and zero otherwise. Since each student only participates a single time, $\mathrm{Technical}_i + \mathrm{Punitive}_i + \mathrm{Shocking}_i = 1$. Students in the control group have $\mathrm{Technical}_i = \mathrm{Punitive}_i = \mathrm{Shocking}_i = 0$, so that the intercept $\alpha$ in (1) absorbs the average drivers' overconfidence of all the sample. The three dummies absorb any linear deviation of the average drivers' overconfidence of that particular treatment group with respect to the

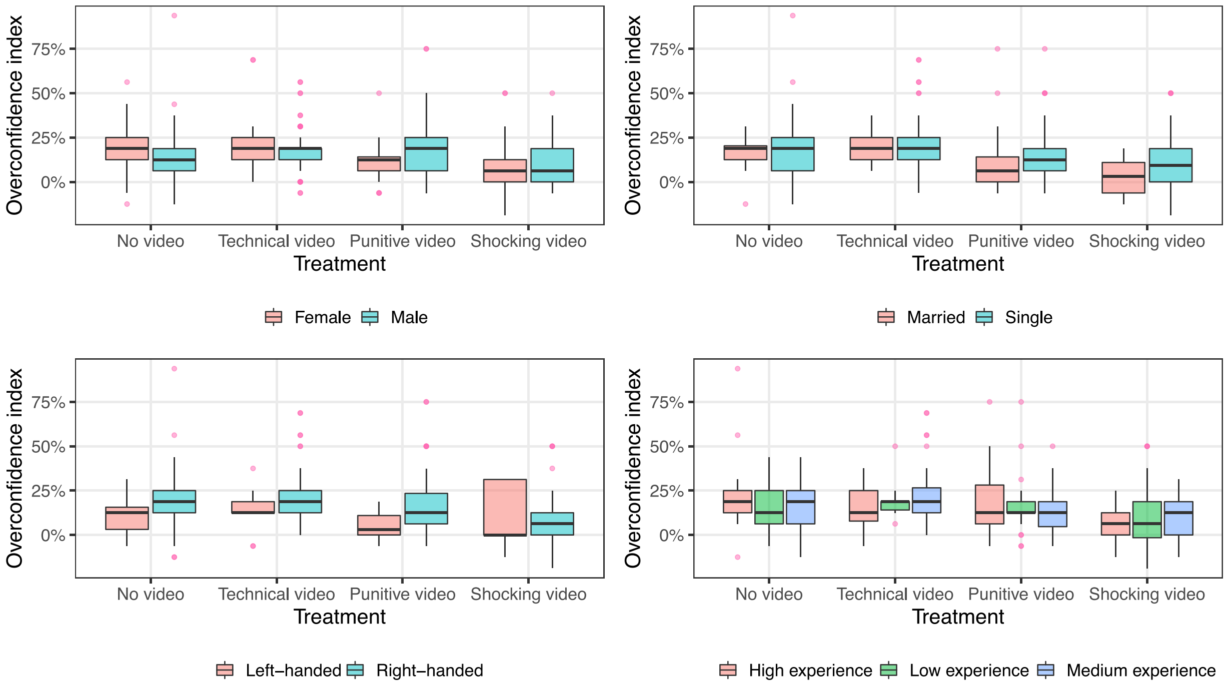


**Fig. 1.** We present boxplots of the overconfidence index for different data subsamples. In the $x$-axis, we segregate by treatment type (no video, technical video, punitive video, and shocking video). The $y$-axis shows the average overconfidence index for each subgroup. Each panel further subdivides the sample into (a) gender (female, male), (b) marital status (married, single), (c) laterality (left- and right-handed), and (d) driving experience (high, medium, and low experience).

entire sample. The term $\epsilon_i$ is the residual, which captures variations of the level of overconfidence that are not explained by the external stimuli nor the average effect.

Since we are looking to purge out the effect of the external intervention and put the drivers' overconfidence level distribution in common grounds for the control and treatment groups, we use the residual $\epsilon_i$ of specification in (1) to identify the most important attributes in explaining students' overconfidence. Such residual gives an unbiased level of overconfidence, which is not contaminated by the external stimuli. The left-most panel of Fig. 2 shows a kernel density plot of the original level of overconfidence of students ($y_i$ in (1)), while the right-most panel shows the same plot but with the unbiased level of overconfidence among students in the control and treatment groups ($\epsilon_i$ in (1)). While the distributions of the level of overconfidence are clearly different, suggesting that the external intervention has an effect on drivers' overconfidence, the distributions become roughly the same if take the residual of specification in (1).

We use an elastic net regression (Zou and Hastie, 2005) to estimate the importance of each attribute in the model. Such regression optimally combines $L_2$-norm (ridge) and $L_1$-norm (Lasso) regularization and is useful when we have several attributes to choose from and few observations. Standard linear models, such as the OLS, would perform poorly as we are in a high-dimensional space[20] with few observations. Regularization is a natural step to overcome this problem, as it balances model's complexity (number of parameters) and performance (). The ridge regularization tends to shrink the coefficients of correlated predictors towards each other while the Lasso tends to pick one of them and discard the others.

To select the most important attributes, we model the unbiased level of overconfidence ($\epsilon_i$) using all attributes we have as follows:

$$\epsilon_i = \boldsymbol{\beta}^T \cdot \mathbf{X}_i + \text{error}_i, \tag{2}$$

in which $\dim(\mathbf{X}) = 400 \times 34$ (400 observations and 33 attributes + 1 constant) and $\dim(\boldsymbol{\beta}) = 34 \times 1$. Following the elastic net procedure, we select $\boldsymbol{\beta}$ that minimizes the following loss function $L(\beta)$:

$$L(\beta) = \sum_{i=1}^{n}\left(\epsilon_i - \sum_{j=1}^{p}\beta_j x_{ij}\right)^2 + \lambda\left[(1-\alpha)\frac{||\boldsymbol{\beta}||_2^2}{2} + \alpha||\boldsymbol{\beta}||_1^2\right], \tag{3}$$

in which $n = 400$ is the number of observations and $p = 33 + 1$ is the number of attributes; $\lambda \geq 0$ is the regularization parameter that penalizes model complexity; and $\alpha \in [0, 1]$ controls the mixture between $L_2$ and $L_1$ regularization. The first term is the traditional error, measured by the projected and the real unbiased overconfidence level. The second is the regularization term, which has relative weight of $\lambda$ with respect to the traditional OLS fitting, and comprises a convex linear combination of $L_2$ regularization (weighted by $\alpha$) and $L_1$ regularization (weighted by $1-\alpha$).

Following the machine learning literature, model complexity is given in terms of the total magnitude of the model free parameters, which are encoded in the vector $\boldsymbol{\beta}$. Like Lasso and ridge, we do not penalize the magnitude of the intercept term, which is one of the coefficients inside $\boldsymbol{\beta}$. When $\lambda = 0$, the estimation of $\boldsymbol{\beta}$ reduces to a standard OLS. If $\lambda \neq 0$, then regularization is enabled. In such case, when $\alpha = 0$, Eq. (3) simplifies to a ridge regression. In contrast, when $\alpha = 1$, Eq. (3) reduces to a Lasso regression.

In the elastic net regression, $\alpha$ takes values in-between 0 and 1. We optimally tune $\alpha$ and $\lambda$ using a nested cross-validation procedure. Such methodology enables us to tune the regularization parameters while preventing overfitting of the model.[21] We use $k = 10$ folds, repeat such process 100 times, and report the average values. We optimize $\alpha$ over

[20] The number of dimensions is equivalent to the number of attributes used in the model. In our case, we have 33 attributes to choose from.

[21] The nested cross-validation procedure involves multiple runs of $k$-fold cross-validation procedures. Each $k$-fold cross-validation estimates an unbiased accuracy for a given set of parameters $\alpha$ and $\lambda$ (inner loop). However, to estimate the optimal $\lambda$ and $\alpha$, we need another $k$-fold cross validation outside the inner $k$-fold cross-validation procedure (outer loop). In each $k$-fold cross-validation procedure, we split the data into $k$ non-overlapping folds or subsets. In each run, one fold is held out for performance check while the model is trained with the $k-1$ remaining subsets. For each pair of parameters $\lambda$ and $\alpha$, we perform this procedure $k$ times, such that each fold is left for performance check exactly once. In the end, we select the parameters $\lambda$ and $\alpha$ that maximize the accuracy of the model.

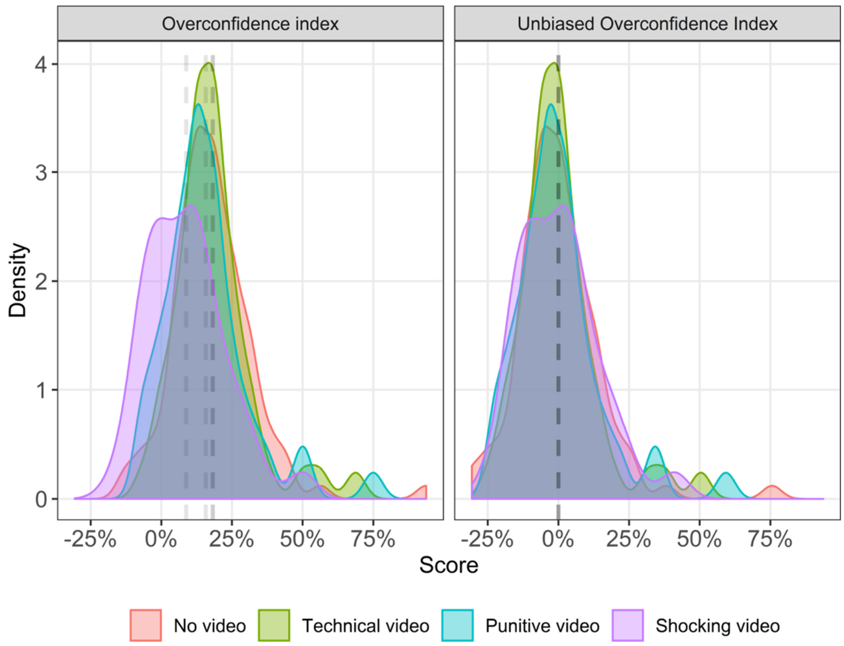


**Fig. 2.** Kernel density plots of (a) the plain level of overconfidence of students and (b) the residual of the level of overconfidence after controlling for the external video stimuli (term $\epsilon_i$ of specification (1)).

the grid search space {0, 0.01, 0.02, ...,1} and $\lambda$ over {0, 1, 2, ...,100}. Also, we pre-process all attributes by applying a $Z$-score standardization, which is a demeaning process followed by a division by the standard deviation of each attribute.

Fig. 3 plots the 14 most relevant (out of 33) attributes for explaining the unbiased level of confidence. The optimal regularization parameters were $\lambda = 2$ and $\alpha = 0.545$. We normalize the coefficients in terms of the most important attribute. The attribute "Drink fines?" is the most powerful predictor of students' level of overconfidence, followed by "Age" and "Cellphone fines." There is an exponential decrease of the importance of attributes, suggesting that few attributes would suffice for the estimation of the level of overconfidence of students. In our econometric specification in the next section, we choose the top 10 attributes.[22]

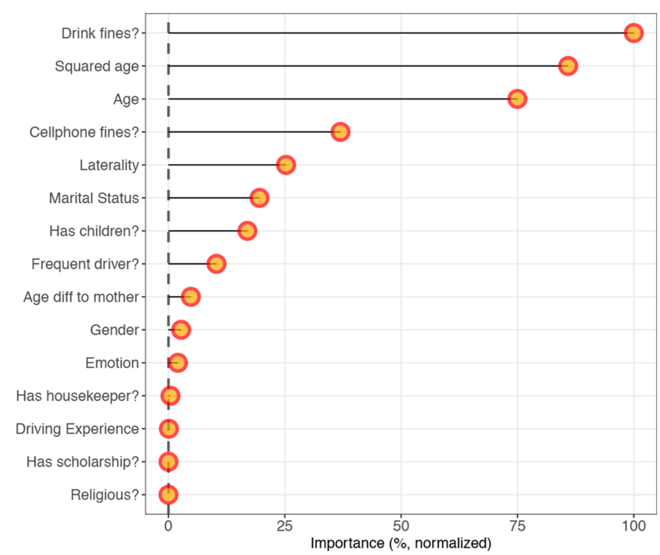


**Fig. 3.** Feature selection results using an elastic net procedure with $L_2$ and $L_1$ regularization. Coefficients are normalized in terms of the most important attribute ("drink fines?").

## 5. Results and discussion

In this section, we define our econometric specifications and report our empirical results.

### 5.1. Measuring the efficiency of different traffic campaign videos in shaping drivers' overconfidence

To test the role of different traffic campaign videos in increasing or reducing drivers' overconfidence, we use the following baseline model:

$$y_{ig} = \alpha_g + \beta_1\,\text{Technical}_i + \beta_2\,\text{Punitive}_i + \beta_3\,\text{Shocking}_i + \gamma^T\,\text{Controls}_i + \epsilon_{ig} \tag{4}$$

in which $i$ indexes students and $g$ groups of students. The variable $y_{ig}$ is the overconfidence level of student $i$ in group $g$; $\alpha_g$ represents fixed effects at different group levels (will be discussed); $\text{Technical}_i$, $\text{Punitive}_i$, $\text{Shocking}_i$ are dummies that indicate whether student $i$ is in the treatment group that watched the technical, punitive, or shocking video, respectively. $\text{Controls}_i$ are control variables that can influence students' overconfidence level, and $\epsilon_{ig}$ is the standard error term. As baseline model, we measure overconfidence index by the average score students obtain in the set of questions in Table 3 (see Section 3.2.2). As controls, we choose the 10 most relevant predictors of students' overconfidence level as identified by our feature extraction methodology (refer to Section 4), which are:

1. *Drink fines?*: dummy variable that equals 1 if the student has received a fine related to alcohol drinking and 0, otherwise.
2. *Age*: numerical variable describing the age of the student.
3. *Cellphone fines?*: dummy variable that equals 1 if the student has received a fine related to cellphone handling and 0, otherwise.
4. *Right-handed*: dummy variable that equals 1 if the student is right-handed and 0, otherwise.
5. *Single*: dummy variable that equals 1 if the student is single and 0, otherwise.
6. *Has children?*: dummy variable that equals 1 if the student has

[22] Observe that the variables below the top 10 approximately do not explain anything of the level of overconfidence of students after controlling for the top 10 attributes.

**Table 3**
Summary statistics of the variables used in the econometric specification. Numerical variables are reported before any logarithmic transformation. The first row corresponds to the dependent variable. The remainder denotes the independent variables, which are ordered by feature importance (see Section 4).

| Statistic | $N$ | Mean | St. Dev. | Min | Pctl(25) | Pctl(75) | Max |
|---|---|---|---|---|---|---|---|
| Overconfidence index | 400 | 0.15 | 0.15 | −0.19 | 0.10 | 0.20 | 1.00 |
| Has drink fines | 400 | 0.04 | 0.21 | 0 | 1 | 1 | 1 |
| Age | 400 | 24.46 | 5.49 | 19 | 22 | 25 | 58 |
| Has cellphone fines | 400 | 0.12 | 0.32 | 0 | 1 | 1 | 1 |
| Right-handed | 400 | 0.90 | 0.30 | 0 | 1 | 1 | 1 |
| Single | 400 | 0.83 | 0.37 | 0 | 1 | 1 | 1 |
| Has children | 400 | 0.17 | 0.37 | 0 | 1 | 1 | 1 |
| Frequent driver | 400 | 0.67 | 0.47 | 0 | 0 | 1 | 1 |
| Mother's–student's age | 400 | 26.30 | 6.44 | 13 | 22 | 30 | 44 |
| Male | 400 | 0.54 | 0.50 | 0 | 0 | 1 | 1 |
| Emotion | 400 | 4.59 | 1.47 | 1 | 4 | 6 | 7 |

**Table 4**
This table reports the output from Regression (4). The dependent variable is the log of the overconfidence index of students plus a constant $k$. We set $k = 1.19$, which is one plus the minimum overconfidence index in absolute values, i.e., $|-0.19|$ according to Table 3. The coefficient estimates of interest are the dummy variables Technical, Punitive, and Shocking videos, which provide elasticities of the efficiency of traffic campaign videos in shaping drivers' overconfidence. We add as controls the ten most predictive variables identified by our machine-learning feature extraction procedure discussed in Section 4. We take the log values of all numerical variables. We use Newey and West (1987)'s robust standard errors and no fixed effects in every specification in this table. Dummies for the three treatment groups are present in every specification. We test different controls from columns (1) to (6) and add all of them in column (7). Statistical significance levels: *** $p$-value <0.01, ** $p$-value <0.05, * $p$-value <0.10.

| Dependent variable | log($k$+ Overconfidence Index) | | | | | | |
|---|---|---|---|---|---|---|---|
| | (1) | (2) | (3) | (4) | (5) | (6) | (7) |
| *Treatment groups* | | | | | | | |
| Technical video | 0.003 | 0.001 | 0.003 | 0.004 | 0.003 | −0.0004 | −0.002 |
| | (0.020) | (0.020) | (0.020) | (0.020) | (0.020) | (0.020) | (0.019) |
| Punitive video | −0.025 | −0.027 | −0.019 | −0.024 | −0.024 | −0.029 | −0.021 |
| | (0.020) | (0.020) | (0.020) | (0.020) | (0.020) | (0.020) | (0.020) |
| Shocking video | −0.100*** | −0.104*** | −0.090*** | −0.097*** | −0.101*** | −0.099*** | −0.089*** |
| | (0.020) | (0.020) | (0.020) | (0.020) | (0.020) | (0.020) | (0.019) |
| *Control variables* | | | | | | | |
| Has drink fines | | 0.114*** | | | | | 0.125*** |
| | | (0.034) | | | | | (0.034) |
| Has cellphone fines | | 0.045** | | | | | 0.031 |
| | | (0.022) | | | | | (0.022) |
| log(age) | | | 1.572** | | | | 1.498** |
| | | | (0.621) | | | | (0.610) |
| $\log^2$(age) | | | −0.259*** | | | | −0.246*** |
| | | | (0.096) | | | | (0.094) |
| Right-handed | | | | 0.055** | | | 0.056** |
| | | | | (0.024) | | | (0.023) |
| Single | | | | | 0.039** | | 0.014 |
| | | | | | (0.019) | | (0.026) |
| Has children | | | | | | −0.045* | −0.040 |
| | | | | | | (0.023) | (0.026) |
| Frequent driver | | | | | | 0.020 | 0.019 |
| | | | | | | (0.015) | (0.015) |
| log(mother's–student's age) | | | | | | 0.006 | 0.026 |
| | | | | | | (0.048) | (0.049) |
| log(Emotion) | | | | | | −0.003 | −0.006 |
| | | | | | | (0.019) | (0.018) |
| Male | | | | | | 0.012 | 0.010 |
| | | | | | | (0.015) | (0.014) |
| Constant | −0.016 | −0.024* | −2.382** | −0.068** | −0.049** | −0.040 | −2.431** |
| | (0.014) | (0.014) | (1.007) | (0.026) | (0.021) | (0.151) | (1.011) |
| Observations | 400 | 400 | 400 | 400 | 400 | 400 | 400 |
| $R^2$ | 0.080 | 0.123 | 0.128 | 0.093 | 0.090 | 0.097 | 0.196 |
| Adjusted $R^2$ | 0.073 | 0.112 | 0.117 | 0.084 | 0.081 | 0.078 | 0.167 |

children and 0, otherwise.

7. *Frequent driver?*: dummy variable that equals 1 is a frequent driver and 0, otherwise.
8. *Age diff to mother*: numerical variable indicating the difference of ages between the mother and the student.
9. *Male*: dummy variable that equals 1 is male and 0, otherwise.
10. *Emotion*: numerical variable that indicates the level of emotion of the student. It ranges from 0 (not emotive at all) to 7 (very emotive).

We estimate (4) using OLS with robust standard errors (Newey and West, 1987). Since we are dealing with a cross-sectional data, such

**Table 5**
This table reports the output from Regression (5). The dependent variable is the log of the overconfidence index of students plus a constant $k$. We set $k = 1.19$, which is one plus the minimum overconfidence index in absolute values, i.e., $|-0.19|$ according to Table 3. The coefficient estimates of interest are the dummy variables Technical, Punitive, and Shocking videos, interacted with male, which capture any heterogeneity among men and women concerning the efficiency of traffic campaign videos in shaping drivers' overconfidence. We add as controls the ten most predictive variables identified by our machine-learning feature extraction procedure discussed in Section 4, except for gender, which is explicitly in the table. We take the log values of all numerical variables. We use Newey and West (1987)'s robust standard errors and no fixed effects in every specification in this table. We gradually add fixed effects that absorb the biological, social, and driving characteristics of students from columns (1) to (6). Statistical significance levels: *** $p$-value <0.01, ** $p$-value <0.05, * $p$-value <0.10.

| *Dependent variable* | log($k$+ Overconfidence Index) | | | | | |
|---|---|---|---|---|---|---|
| | (1) | (2) | (3) | (4) | (5) | (6) |
| *Interactions (treatment × gender)* | | | | | | |
| Technical video × Male? | 0.016 | 0.014 | 0.002 | 0.006 | 0.008 | −0.011 |
| | (0.041) | (0.041) | (0.043) | (0.041) | (0.041) | (0.046) |
| Punitive video × Male? | 0.098** | 0.100** | 0.089** | 0.087** | 0.082** | 0.077* |
| | (0.042) | (0.041) | (0.043) | (0.041) | (0.041) | (0.045) |
| Shocking video × Male? | 0.049 | 0.060 | 0.052 | 0.051 | 0.047 | 0.030 |
| | (0.040) | (0.040) | (0.042) | (0.040) | (0.040) | (0.045) |
| *Non-interacted variables (treatment + gender)* | | | | | | |
| Technical video | −0.0001 | −0.002 | 0.013 | 0.011 | 0.010 | 0.012 |
| | (0.030) | (0.030) | (0.031) | (0.030) | (0.030) | (0.032) |
| Punitive video | −0.084*** | −0.086*** | −0.074** | −0.073** | −0.070** | −0.089*** |
| | (0.031) | (0.030) | (0.032) | (0.031) | (0.031) | (0.033) |
| Shocking video | −0.121*** | −0.127*** | −0.107*** | −0.106*** | −0.102*** | −0.093*** |
| | (0.026) | (0.026) | (0.027) | (0.026) | (0.026) | (0.028) |
| Male | −0.030 | −0.034 | −0.026 | −0.025 | −0.022 | −0.021 |
| | (0.029) | (0.028) | (0.030) | (0.028) | (0.028) | (0.032) |
| Constant | −0.004 | | | | | |
| | (0.018) | | | | | |
| *Fixed effects* | | | | | | |
| Drink fines? | No | Yes | Yes | Yes | Yes | Yes |
| Age + squared Age | No | No | Yes | Yes | Yes | Yes |
| Cellphone fines? | No | No | No | Yes | Yes | Yes |
| Laterality | No | No | No | No | Yes | Yes |
| Marital Status | No | No | No | No | No | Yes |
| Has children | No | No | No | No | No | Yes |
| Frequent driver | No | No | No | No | No | Yes |
| Mother's–student's age | No | No | No | No | No | Yes |
| Emotion | No | No | No | No | No | Yes |
| Observations | 400 | 400 | 400 | 400 | 400 | 400 |
| $R^2$ | 0.096 | 0.131 | 0.269 | 0.271 | 0.286 | 0.385 |
| Adjusted $R^2$ | 0.080 | 0.113 | 0.113 | 0.188 | 0.198 | 0.160 |

error clustering minimizes heteroskedasticity issues. In addition, if variable $x$ is numerical, then we apply the following logarithmic transformation: $\log(k + x)$, in which $k = 1 + |\min(x)|$.[23]

Table 3 reports summary statistics of the overconfidence index and the regressors employed in the model. Even though the overconfidence ranges from −2 to 2, the minimum value is −0.19, and the maximum is 1, suggesting we do not find extreme overconfidence traits in our sample. Also, the mean and standard values of the overconfidence are $0.15 \pm 0.15$, suggesting that students are, on average, overconfident. Only 4% of respondents received a fine related to alcohol driving before driving and 12% with fines related to cellphone handling during driving. We have a wide range of students' age, going from 19 to 58 with a mean value of 24.46. 90% of students are right-handed, 83% are single, 17% have children, 67% are frequent drivers, and 54% are men. In the questionnaire, we also ask students about their current emotion levels, which can range from 1 – not emotive at all to 7 – incredibly emotive. We have a substantial heterogeneity of emotion levels, including students with minimum and maximum emotion levels. The average emotion level of students is 4.59, indicating a degree of emotiveness of students.

Table 4 presents the results from estimating Eq. (4). In the seven econometric specifications, columns (1)–(7), we always introduce the treatment group dummies to test whether traffic campaign videos with the same content but with a different exhibition manner matter. From columns (1) to (6), we add, in an isolated way, controls. In column (7), we introduce all controls at once.

Comparatively to the control group, shocking videos lead to a decrease of 8.9–10.1% in respondents' level of overconfidence. The coefficient does not change much as we add more controls, suggesting that the result is robust. In contrast, the level of overconfidence of students that experienced punitive videos is 2.1–2.9% less than that of the control group, but not statistically significant when we cluster with robust standard errors. The coefficient measuring the relative change in overconfidence of students that watched technical videos comparatively to the baseline group is very close to zero and statistically insignificant. Our results indicate that videos of the Australian school (shocking videos), with strong and life-threatening scenes, are more effective in reducing drivers' overconfidence than videos of the American school (punitive videos), with punitive contents showing the consequences and penalties of drinking and driving, and videos of the European school (technical videos), with technical content.

### *5.2. Does the efficiency of different traffic campaign videos differ across genders?*

We now analyze whether gender plays a role in the efficiency of different traffic campaign strategies. Our baseline econometric

[23] The addition of $k$ prevents the application of the logarithmic transformation to a non-positive number (undefined). We also add 1 to prevent the application of the logarithm to a zero number if the domain of $x$ contains a zero.

specification stays the same, but now we include an interaction of gender with the treatment variables to capture any potential heterogeneity among men and women. Therefore, we use the following model:

$$y_{\mathrm{ig}} = \alpha_g + \sum_{\mathrm{treat}\in\mathcal{T}} \beta_{\mathrm{treat}} \cdot \mathbb{1}_{\{i\,\in\,\mathrm{treat}\}} + \sum_{\mathrm{treat}\in\mathcal{T}} \gamma_{\mathrm{treat}} \cdot \mathbb{1}_{\{i\,\in\,\mathrm{treat}\}} \times \mathrm{Male}_i + \lambda\,\mathrm{Male}_i + \epsilon_{\mathrm{ig}} \tag{5}$$

in which $i$ indexes students and $g$ groups of students; $y_{\mathrm{ig}}$ is the overconfidence level of student $i$, member of the group $g$; and the other variables follow the same convention as the previous econometric exercises. Our coefficient of interest is $\gamma$, which measures any potential heterogeneity in men's and women's overconfidence levels after watching the same traffic video content but with different exhibition strategies. We also add the marginal term $\mathrm{Male}_i$ to capture the average overconfidence level of men that differs from the average overconfidence level of the entire sample (men + women).

Table 5 reports the results of Regression (5). Columns (1)–(6) incrementally add fixed effects that capture biological, social and driving heterogeneities among students. Looking at the interacted terms, we observe that punitive videos are less effective in reducing the overconfidence of men. Women that watched the punitive video have 7.0–8.9% less overconfidence that the control group. However, the overall effect of punitive video on men's overconfidence ranges from −1.2% to +1.5%, with no statistically significance (compound effect).

Our findings indicate that shocking videos are an excellent traffic campaign strategy to reduce overconfidence levels regardless of gender. Punitive videos, in contrast, are valid only for women, being ineffective for men. Finally, we find no empirical evidence that technical videos change students' overconfidence.

## 6. Conclusions

Promoting traffic safety is one of the most important goals for public policymakers in today's society and represents a critical strategic issue to reduce the number of traffic accidents, as according to data from the World Health Organization (WHO) road accidents kill 1.25 million people a year worldwide and are the leading cause of death for people aged 15–29. In Brazil, according to DataSUS, more than 37.3 thousand people die every year in the transit of cities and highways of the country. A significant number of these deaths is caused due to the Handling the cell phone, and the consumption of alcoholic beverage in the direction is generating a substantial amount of deaths.

In this way, pursuing effective public policies to reduce these numbers is of fundamental importance. In this sense, this work stresses the importance of paying attention to the behavior of drivers in traffic, since they are responsible for transgressions and deaths in transit. This work, seeking a broad look at the practice of drivers in traffic, reveals that their behavior can be affected by mood swings. Punitive and strong-minded traffic campaigns have proved effective in reducing drivers' overconfidence, thus providing more conscious and cautious action in traffic. Also, this work sought to trace the individual characteristics of drivers that influence their perception of risk and the construction of their overconfidence.

Because it is such a current and recurring issue, future research may try to explain these behaviors through field experiments, testing in practice what influences individuals to commit such traffic offenses, and thus contributing to the formulation of public policies.

We have several contributions in our paper. We show how to use machine learning techniques to explore more robust regressions, controlling for many factors as possible. We find that campaigns matter to reduce overconfidence, and the message is relevant. Further research could use a similar approach and test the short versus long-term effectiveness of safe driving campaigns. Increasing the sample to have a more heterogeneous sample with regards to age may also provide further insights, as some campaigns may be more effective in specific age groups.

## Declaration of Competing Interest

The authors have no conflicts of interest to declare.

## Appendix A. Supplementary data

Supplementary data associated with this article can be found, in the online version, at https://doi.org/10.1016/j.aap.2020.105694.

## References

Agrawal, N., Duhachek, A., 2010. Emotional compatibility and the effectiveness of antidrinking messages: a defensive processing perspective on shame and guilt. J. Market. Res. 47, 263–273.
Atchley, P., Atwood, S., Boulton, A., 2011. The choice to text and drive in younger drivers: behavior may shape attitude. Accid. Anal. Prev. 43, 134–142.
Bayer, J., Campbell, S., 2012. Texting while driving on automatic: considering the frequency-independent side of habit. Comput. Human Behav. 28, 2083–2090.
Bick, T., Hohenemser, C., 1980. Target: highway risks. Environment 21, 17–40.
Bucchi, A., Sangiorgi, C., Vignali, V., 2012. Traffic psychology and driver behavior. Proc. Soc. Behav. Sci. 53, 972–979.
Cameron, M., Haworth, N., Oxley, J., Newstead, S., Le, T., 1993. Evaluation of Transport Accident Commission Road Safety Television Advertising. Technical Report Monash University Accident Research Centre Victoria.
Cameron, M., Vulcan, P., 1998. Evaluation Review of the Supplementary Road Safety Package and Its Outcomes During the First Two Years. Technical Report Camcomp Partners and Research and Safety Services Wellington, New Zealand.
Charness, G., Gneezy, U., Kuhn, M.A., 2012. Experimental methods: between-subject and within-subject design. J. Econ. Behav. Organ. 81, 1–8.
Chen, Y., Lu, F., Zhang, J., 2017. Social comparisons, status and driving behavior. J. Public Econ. 155, 11–20.
Cohen, J., 1960. Chance, Skill and Luck: The Psychology of Guessing and Gambling. Oxford, England.
Cohen, J., 1988. Statistical Power Analysis for the Behavioral Sciences. Erlbaum, Hillsdale, NJ.
de Craen, S., Twisk, D.A., Hagenzieker, M.P., Elffers, H., Brookhuis, K.A., 2008. The development of a method to measure speed adaptation to traffic complexity: identifying novice, unsafe, and overconfident drivers. Accid. Anal. Prev. 40, 1524–1530.
Ehsani, J., Bingham, C., Ionides, E., Childers, D., 2014. The impact of Michigan's text messaging restriction on motor vehicle crashes. J. Adolesc. Health 54, S68–S74.
Faul, F., Erdfelder, E., Lang, A., Buchner, A., 2007. G*Power 3: a flexible statistical power analysis program for the social, behavioral, and biomedical sciences. Behav. Res. Methods 39, 175–191.
Freeman, J., Szogi, E., Truelove, V., Vingilis, E., 2016. The law isn't everything: The impact of legal and non-legal sanctions on motorists' drink driving behaviors. J. Saf. Res. 59, 53–60.
Goodwin, A., O'Brien, N., Foss, R., 2012. Effect of North Carolina's restriction on teenage driver cell phone use two years after implementation. Accid. Anal. Prev. 48, 363–367.
Gross, J.J., Levenson, R.W., 1995. Emotion elicitation using films. Cogn. Emot. 9, 87–108.
Hayashi, Y., Miller, K., Foreman, A., Wirth, O., 2016. A behavioral economic analysis of texting while driving: delay discounting processes. Accid. Anal. Prev. 97, 132–140.
Ifcher, J., Zarghamee, H., 2011. Happiness and time preference: the effect of positive affect in a random-assignment experiment. Am. Econ. Rev. 101, 3109–3129.
Ifcher, J., Zarghamee, H., 2014. Affect and overconfidence: a laboratory investigation. J. Neurosci. Psychol. Econ. 7, 125–150.
Isen, A.M., 2007. Positive affect, cognitive flexibility, and self-control. In: Shoda, Y., Cervone, D., Downey, G. (Eds.), Persons in Context: Building a Science of the Individual. The Guilford Press, New York, pp. 130–147.
Isen, A.M., 2008. Some ways in which positive affect influences decision making and problem solving. In: Lewis, M., Haviland-Jones, J.M., Barrett, L.F. (Eds.), Handbook of Emotions. The Guilford Press, New York, pp. 548–573.
Johansson, G., Rumar, K., 1968. Visible distances and safe approach speeds for night driving. Ergonomics 11, 275–282.
Karama, S., Lecours, A., Leroux, J., Bourgouin, P., Beaudoin, G., Joubert, S., Beauregard, M., 2002. Ares of brain activation in males and females during viewing of erotic film exerpts. Human Brain Mapp. 16, 1–13.
Kirchsteiger, G., Rigotti, L., Rustichini, A., 2006. Your morals might be your moods. J. Econ. Behav. Organ. 59, 155–172.
List, J., Sadoff, S., Wagner, M., 2011. So you want to run an experiment, now what? Some simple rules of thumb for optimal experimental design. Exp. Econ. 14, 439–457.
Lyubomirsky, S., King, L., Diener, E., 2005. The benefits of frequent positive affect: Does happiness lead to success? Psychol. Bull. 131, 803–855.
Martin, M., 1990. On the induction of mood. Clin. Psychol. Rev. 10, 669–697.
Matthews, M.L., Moran, A.R., 1986. Age differences in male drivers' perception of accident risk: the role of perceived driving ability. Accid. Anal. Prev. 18, 299–313.
Merkle, C., Weber, M., 2011. True overconfidence: the inability of rational information processing to account for apparent overconfidence. Organ. Behav. Human Decis.

Process. 116, 262–271.
Murry, J., Stam, A., Lastovicka, J., 1993. Evaluating an anti-drinking and driving advertising campaign with a sample survey and time series intervention analysis. J. Am. Stat. Assoc. 421, 50–56.
Newey, W., West, K.D., 1987. A simple, positive semi-definite, heteroskedasticity and auto-correlation consistent-covariance matrix. Econometrica 55, 703–708.
Newstead, S., Cameron, M., Gantzer, S., Vulcan, P., 1995. Modelling of Some Major Factors Influencing Road Trauma Trends in Victoria in 1989–93. Technical Report 74 Monash University Accident Research Centre Victoria.
Palomba, D., Sarlo, M., Angrilli, A., Mini, A., Stegagno, L., 2000. Cardiac responses associated with affective processing of unpleasent film stimuli. Int. J. Psychophysiol. 36, 45–57.
Pannenberg, M., Friehe, T., 2019. Does it really get better with age? Life-cycle patterns of confidence in Germany. Annual Conference 2019 (Leipzig): 30 Years after the Fall of the Berlin Wall – Democracy and Market Economy Verein für Socialpolitik /German Economic Association.
Philippot, P., 1993. Inducing and assessing differentiated emotion feeling states in the laboratory. Cogn. Emot. 1, 171–193.
Phillips, R.O., Ulleberg, P., Vaa, T., 2011. Meta-analysis of the effect of road safety campaigns on accidents. Accid. Anal. Prev. 43, 1204–1218.
Quisenberry, P., 2015. Texting and driving: can it be explained by the general theory of crime? Am. J. Criminal Just. 40, 303–316.
Rimal, R.N., Yilma, H., Ryskulova, N., Geber, S., 2019. Driven to succeed: improving adolescents' driving behaviors through a personal narrative-based psychosocial intervention in Serbia. Accid. Anal. Prev. 122, 172–180.
Rottenberg, J., Ray, R.D., Gross, J.J., 2007. Emotion elicitation using films. In: Coan, J.A., Allen, J.J.B. (Eds.), The Handbook of Emotion Elicitation and Assessment. Oxford University Press, London, pp. 9–28.
Shinar, D., Rockwell, T., Malecki, J., 1980. The effects of changes in driver perception on rural curve negotiation. Ergonomics 23, 263–275.
Struckman-Johnson, C., Gaster, S., Struckman-Johnson, D., Johnson, M., May-Shinagle, G., 2015. Gender differences in psychosocial predictors of texting while driving. Accid. Anal. Prev. 74, 218–228.
Summala, H., 1987. Young driver accidents: risk taking or failure of skills? Alcohol Drugs Driving 3, 79–91.
Svenson, O., 1970. A functional measurement approach to intuitive estimation as exemplified by estimated time savings. J. Exp. Psychol. 86, 204–210.
Svenson, O., 1978. Risks of road transportation in a psychological perspective. Accid. Anal. Prev. 10, 267–280.
Svenson, O., 1981. Are we all less risky and more skillful than our fellow drivers? Acta Psychol. 47, 143–148.
Svenson, O., Fischhoff, B., MacGregor, D., 1985. Perceived driving safety and seatbelt usage. Accid. Anal. Prev. 17, 119–164.
SWOV, 2009. SWOV Fact Sheet: Fear-Based Information Campaigns. Technical Report Institute for Road Safety Research Leidschendam, the Netherlands.
Tucker, S., Pek, S., Morrish, J., Ruf, M., 2015. Prevalence of texting while driving and other risky driving behaviors among young people in Ontario, Canada: Evidence from 2012 and 2014. Accid. Anal. Prev. 84, 144–152.
Warmath, D., Piehlmaier, D., Robb, C., 2019. The impact of shared financial decision making on overconfidence for married adults. Financ. Plann. Rev. 2, e1032.
Weinstein, N., 1980. Unrealistic optimism about future life events. J. Pers. Soc. Psychol. 39, 806–820.
Westerman, R., Spies, K., Stahl, G., Hesse, F., 1996. Relative effectiveness and validity of mood induction procedures. Eur. J. Soc. Psychol. 26, 557–580.
WHO, 2013. Global Status Report on Road Safety 2013.
WHO, 2018. Global Status Report on Road Safety 2018.
Wohleber, R.W., Matthews, G., 2016. Multiple facets of overconfidence: Implications for driving safety. Transp. Res. Part F: Traff. Psychol. Behav. 43, 265–278.
Zou, H., Hastie, T., 2005. Regularization and variable selection via the elastic net. J. R. Stat. Soc.: Ser. B Stat. Methodol. 67, 301–320.