Can information really change travel behaviour? Controlling for endogeneity in modelling the effect of feedback information

David Palma [*], Romain Crastes Dit Sourd,
Chiara Calastri, Stephane Hess
Choice Modelling Centre and Institute for Transport Studies
University of Leeds

Matthew J. Beck
Institute of Transport and Logistic Studies, Business School
The University of Sydney

Vikki O'Neill
Centre for Public Health, School of Medicine,
Dentistry and Biomedical Sciences
Queen's University

2019
April

Abstract

As congestion and pollution increase in our cities, there is much interest in cost-effective ways to encourage use of active and public modes of transport. We tested the efficacy of passively providing information to travellers to persuade them to increase their use of such modes. We tracked the travel behaviour of 454 individuals for two weeks. At the beginning of the second week, we split them into three groups receiving either: (i) no information, (ii) information about their own travel behaviour, or (iii) information comparing their travel behaviour to that of other similar participants. We analysed the data using a difference-in-difference approach, correcting for the endogeneity of information type (iii) using the 2-stage least square approach. Our results, unlike other studies in the literature, reveal no significant effect of providing information to participants. While this could be due to our experimental settings, we believe previous positive results could have been due to ignoring the influence of endogeneity on comparative information. Indeed, participants' unobserved characteristics (e.g. being sedentary) could influence both their original use of active modes as well as their reaction to the information, and controlling for this shows that the impact of the information provision is negligible.

Keywords: information provision; feedback; travel behaviour; smartphone tracking

[*]D.Palma@leeds.ac.uk

1 Introduction

Transport planners and researchers are looking for ways to incentivise the use of active travel modes and reduce the use of car. Increased walking and cycling not only reduces traffic, but also generates health benefits among the individuals that increase their physical activity, and to the city as a whole, due to reduced pollution and a more friendly urban landscape (de Nazelle et al., 2011).

While provision of new infrastructure is known to incentivise modal change (Pucher et al., 2010), it is usually expensive, requires long period of planning and construction, and can cause relevant disruptions to the urban (and non-urban) environment. Looking for more cost-effective alternatives, and probably inspired by the development of behavioural economics (Thaler, 1980), researchers have turned their attention towards persuading travellers into changing their behaviour. Such an approach would cost a fraction of what infrastructure changes would, and could potentially lead to more efficient uses of current resources.

At the same time, information and communication technologies (ICT) -in particular the smartphone- has revealed themselves as a powerful and cost-effective way to both collect information about individuals (Wang et al., 2018) and provide new information to them (Weiser et al., 2016). As smartphone use is strongly engrained in people's daily life (Oulasvirta et al., 2012), it is seen as an affordable way to reach a massive audience. Accordingly, a growing body of literature has developed around the most efficient way to persuade travellers into choosing active travel modes, by the means of *app* (*i.e.* software) accessible through their own smartphones.

However, persuasion through the use of ICT is not straightforward. Its effectiveness depends on a myriad of variables, and is highly dependent on the kind of stimulus used, as well as on the context (Fogg, 2002). In the specific setting of travel information, Chorus et al. (2006) points to the low willingness to pay for information through ICTs, and how familiarity with internet and personal devises is paramount to use it. Chorus et al. (2013) propose a model for information acquisition based on the expected utility of that information for a given travel choice, implying that the relevance and impact of information will be determined mainly by the context. Ben-Elia and Avineri (2015) also highlight the relevance of both the stimulus as well as the context to understand the persuasiveness of information provided through ICTs to travellers. They also point to a lack of understanding of how suggestions are adopted or rejected by travellers, and how sensitivity to information decays as its provision becomes familiar. In the specific context of smartphone apps, Ettema (2018) discusses how the intentions of both software providers and users must match for the tool to be effective.

According to the classification by Anagnostopoulou et al. (2017), which is based on Orji et al. (2014), there are eight basic strategies to persuade individuals to change their travel behaviour through the use of smartphone apps.

- *Comparison* is the strategy where the individual's behaviour is compared to that of others.

- *Self-monitoring* consists in presenting individuals with metrics related to their own travel behaviour.

- *Suggestion* consists in suggesting more environmentally-friendly alternatives to the individual (e.g. the use of public transport instead of car).

- *Simulation* consists in graphically representing the consequences (*i.e.* negative externalities) of the alternatives the user could choose.

- *Cooperation* consists in an approach where a group of individuals cooperate to reach a pre-defined goal (e.g. cycling an aggregate of 100 km during a month).

- *Praise* involves providing positive feedback to travellers when they achieve a pre-defined goal.

- *Personalization* requires presenting suggestions and recommendations to travellers that are specifically tailored to them, based on their past behaviour.

- *Competition* requires setting goals that must be pursued individually, but whose achievement can be seen by others.

Many of these strategies can be used simultaneously, and several can be classified within the larger *gamification* concept. Gamification is the process of "applying game mechanics to non-game contexts in order to engage audiences and to inject a little fun into mundane activities besides generating motivational and cognitive benefits" (Sardi et al., 2017).

A number of studies has looked at the development and efficacy of different persuasion strategies. The most common approach seems to be *praise* (Barrat, 2017; Di Dio et al., 2018; Piras et al., 2018; Weber et al., 2018). *Suggestion* is another common strategy (Bucher et al., 2019; Meloni et al., 2013; Piras et al., 2018). *Competition* is used by Barrat (2017), Di Dio et al. (2018), and Weber et al. (2018). Bucher et al. (2019) is one of the few using *simulation*, while Weber et al. (2018) uses *cooperation*. With the only exception of Bucher et al. (2019), who only does a theoretical impact analysis, all of these studies report a significant impact on travel behaviour due to the use of their respective interventions. Jariyasunant et al. (2015) use the *comparison* persuasion strategy. In this study, the authors develop a smartphone app that tracks participants and provides them with information on their travel behaviour, comparing it to the behaviour of others. The authors find a significant reduction of driving and a small increase on the use of active travel modes due to the intervention.

Most of the existing studies, though, have a number of methodological limitations when it comes to measuring the impact of their interventions. Firstly, several studies are based on stated preferences surveys (Kim et al., 2014; Kuwano et al., 2014; Rasouli and Timmermans, 2016). Even in general applications, stated preference surveys come with a risk of hypothetical bias (Murphy et al., 2005), but this is likely to be exacerbated when looking at the effect of *comparison* e.g. the share of friends conducting specific activities (Norwood and Lusk, 2011). Secondly, most studies only focus on the discrete choice, e.g. mode choice, but not on a continuous aspect of it, e.g. distance travelled by each mode. Another common limitation among studies measuring the impact of the *comparison* persuasion strategy is that often there is no other kind of information provided. In these scenarios, it is not possible to distinguish between the effect of receiving any information at all, and the effect of *comparison* information. The confounding effect is furthered by the fact that some studies do not feature a control group where no information is provided to participants.

Finally, most studies do not account for the endogeneity of all *comparison* feedback. As the information provided is ultimately the difference between one's own behaviour and that of others, that information is inevitably correlated with the individual's unobservable characteristics, and therefore it is endogenous to the individuals' response. For example, an individual with a very active lifestyle will be more likely to walk more than the average individual, and at the same time will be more likely to reduce her use of car due to the information provided, as she is more willing to use active travel modes. Ignoring this issue can lead to bias in model estimation (Guevara, 2015).

Our research aims to measure the efficacy of information provision through a smartphone app on increasing the use of active travel and reducing the use of car. We compare the effect of providing two types of information to individuals, inspired by the *self-monitoring* and *comparison* persuasion strategies. We study the effect of these interventions employing revealed preference data, explicitly providing information about the behaviour of peers, using an adequate control group and correcting for endogeneity. In particular, we study the effect of feedback information on the distance travelled by different modes. We focus on distance because it strongly correlates with CO_2 emissions and calories burned while travelling, two important variables from a public policy perspective. We could have chosen other measures of travel behaviour, but they all have disadvantages as compared to the distance travelled. For example, the number of trips performed in each mode does not correlate as strongly with CO_2 emissions as travelled distance does (Mokhtarian and Varma, 1998), neither does time spent travelling in each mode, as revealed by emissions rates being associated to travelled distance and not time (European Environment Agency, 2016).

The remainder of this paper is organised as follows. The next section presents the data collection, including a description of the feedback provided and the sample. Section 3 discusses the modelling methodology employed. This is followed by the results of the analysis in Section 4 and their discussion in Section 5.

2 Survey work

Our study measures the effect of information provision by comparing travellers' behaviour before and after receiving information. Travel behaviour is measured via a smartphone tracking *app* that records all trips made by participants. The data used in this study was collected within a larger data collection effort described in detail in Calastri et al. (2018). In this section, we only focus on the parts relevant to this study. The following four subsections provide details about each stage of the data collection, the information provided to participants, their main characteristics, and their travel behaviour.

2.1 Procedure

Figure 1 displays a graphical representation of the data collection structure. Data was collected in three stages: (i) completion of an online questionnaire, where participants provided their socio-demographic characteristics and answered a series of attitudinal questions; (ii) first week of tracking via the smartphone *app* and (iii) second week of tracking. In the first stage, participants were asked to complete an online survey describing their socio-demographic characteristics, current dwelling, and a short questionnaire about their attitudes towards social influence. Participants also completed a life-course calendar (a list of important life events) and a name generator (a list of friends and family). From this stage, only the socio-demographics and the attitudinal data is used in this study. The attitudinal data was composed of a novel ten-item questionnaire dealing with individual's susceptibility to interpersonal influence, particularly that related to exercise habits and environmentally friendly behaviour. While original, these statements were based on the underlying ideas of the *Susceptibility to interpersonal influence scale* (Bearden et al., 1989).

The second stage of data collection involved recording the travel behaviour of participants for one week. To do this, participants were asked to install the *rMove* mobility *app* on their smartphones (Resource Systems Group, 2017), which tracked their movements through Global Positioning System (GPS). Every time the application detected that a trip had finished, it prompted participants to specify their trip purpose and destination, mode used to get there, cost (if motorised) of the journey, and third parties (from the name generator or others) involved in the trip/activity.

After the first week of tracking, respondents were randomly assigned to one of three feedback information groups: *control*, *self-monitoring*, or *comparison*. Table 1 summarises the number of participants in each feedback group. The *control* group (261 individuals) did not receive any information at the end of the first week. The *self-monitoring* group (108 individuals) received a short report about their own travel behaviour and out-of-home activities. The *comparison* group (84 individuals) received a report not only about their own travel behaviour activities, but also about the average behaviour of other similar participants during their first week. Individuals in

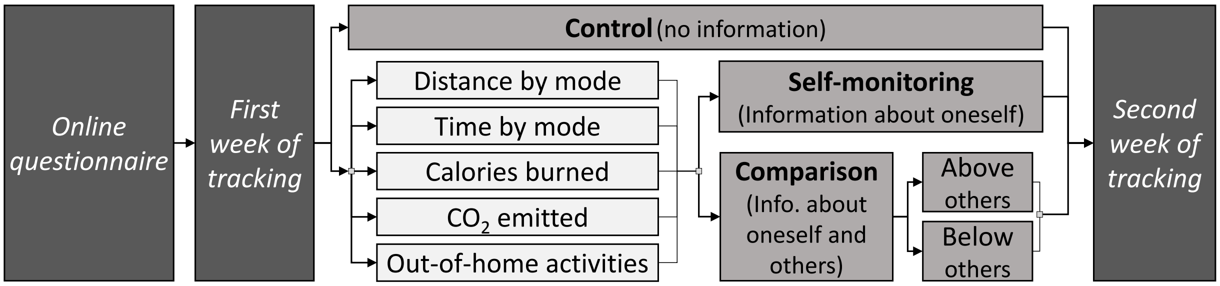


Figure 1: Participants were split in three groups: control, self-monitoring, and comparison feedback

the *comparison* group could find themselves above or below the average behaviour of others for each reported aspect, e.g. a participant could have walked for longer than others, but driven shorter distances than others like him/her. Therefore, there is not a single *above* group, but as many as reported types of behaviours: *above walking distance*, *above walking time*, *above driving distance*, ..., *above CO_2 emissions*, etc. The same is true for the *below* groups. Table 2 presents the number of observations for each subgroup within the *comparison* feedback group.

The third stage of data collection involved monitoring participants behaviour via *rMove* for a second week, just as it was done the week before.

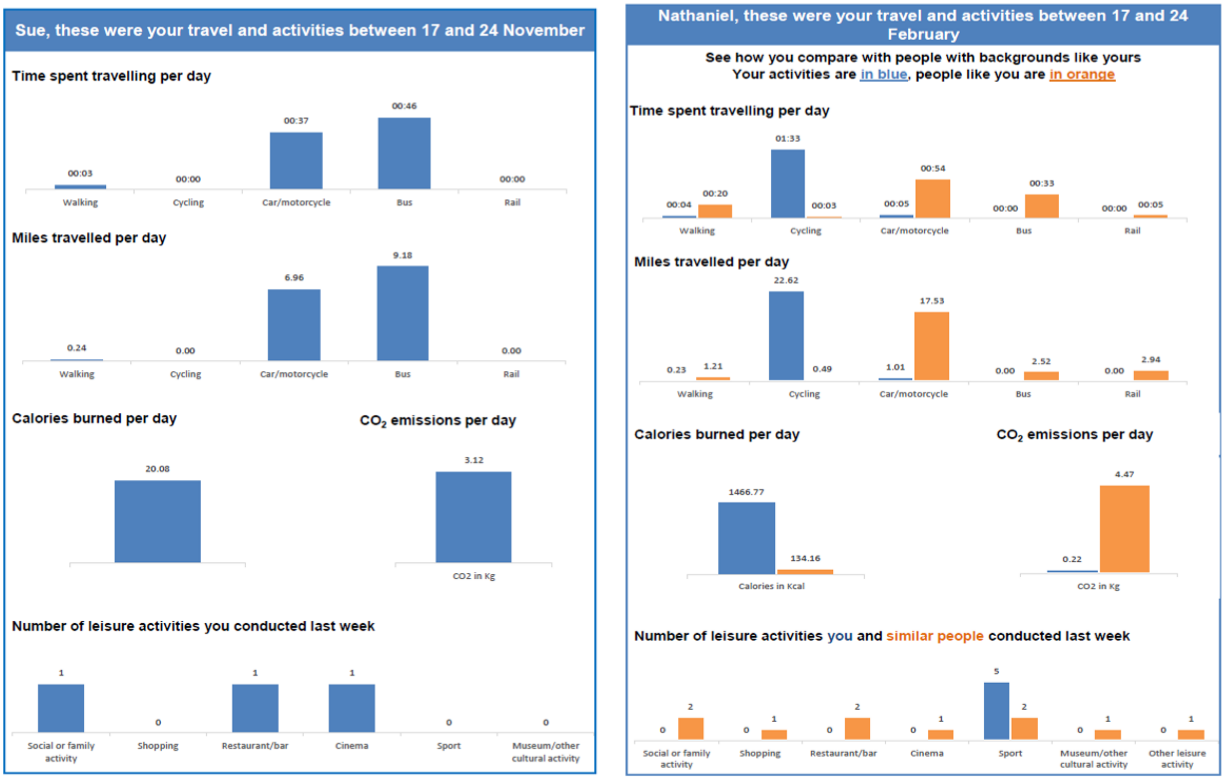
Table 1: Number of participants and observed days in each feedback group

	Observed days	Individuals
Control	1694	261
Self-monitoring	567	108
Comparison	454	84
Total	2715	453

Table 2: Number of observed days in each *comparison* feedback subgroup

	Distance				Time				Calories	CO_2
	Walk	Rail	Bus	Car	Walk	Rail	Bus	Car		
Above	121	106	128	153	157	143	135	198	93	155
Below	333	348	326	301	297	311	319	256	361	293
Total	454	454	454	454	454	454	454	454	454	448*

** One individual did not receive feedback on CO_2*



(a) *Self-monitoring* (b) *Comparison*

Figure 2: Two examples of feedback reports, as seen by participants in the *self-monitoring* (A) and *comparison* (B) groups.

2.2 Feedback information

Figure 2 presents two examples of feedback reports, as shown to participants in the *self-monitoring* and *comparison* groups. They included information on 18 different aspects of travel behaviour, which we call *response variables*. In particular, the report presented the daily average distance travelled and time spent travelling by each mode, as well as the daily average number of calories burned and kilograms (Kg) of CO_2 emitted due to travelling. It also included the number and type of leisure out-of-home activities during the first week. Five modes of transportation were considered: walking, cycling, private motorised vehicles (most notably car, but it also included motorcycle as well as taxi), bus, and rail. The out-of-home activities were classified into social or family activities, shopping, going to a restaurant or bar, going to the cinema or other night out, sports activities, visiting a museum or other cultural activity.

The feedback was e-mailed to participants in the *self-monitoring* and *comparison* groups as a pdf file, with an electronic flag system recording the time when the participant downloaded and looked at the file. Not all participants looked at their feedback immediately after receiving it, but regardless of this, tracking stopped after 14 days for most participants (a few provided additional days of data). Therefore, individuals were observed for a different number of days after reading their feedback. For example, if a participant's tracking begun on a Monday and she received her feedback next Sunday, but looked at it two days later on Tuesday evening, then we would observe the influence of the feedback only for five days (Wednesday to Sunday of the second week of tracking). The first day of observations, *i.e.* the day the participant installed the application, and the day the participant saw the feedback, were excluded from the analysis to avoid potential bias. Such bias could arise, for example, if the participant installed the tracking *app* late on the first day, or looked at the feedback on the late evening, or made more mistakes using the *app* during the first day than later on.

Participants were assigned to a feedback group at the end of the first week of tracking (rather than at the outset of the study) to avoid bias due to attrition between the first stage and the end of the first week of tracking. This way, our group assignment only considered individuals that had shown commitment to the study, reducing the chances of one group being slimmer due to being randomly assigned less committed individuals at the beginning of the study. The size of the *self-monitoring* and *comparison* feedback groups were limited to 120 and 100 individuals, respectively, to keep a larger number of observations free of any feedback for future analysis. However, the final amount of individuals in these groups were smaller than 100 due to attrition during the second week.

The first wave of participants (*i.e.* those recruited up to the first week of November 2016), were only assigned to the *control* or *self-monitoring* groups. They were not assigned to the *comparison* group because there was not enough data to perform an appropriate comparison between participants. While we are aware this breaks the fully randomised assignment, assignment was still random as participants could be assigned to either of two groups. Also, this did not lead to any observable systematic difference in participants' socio-demographics between groups, as we discuss in detail in Section 2.3.

While assignment of participants to the *control*, *self-monitoring* and *comparison* groups was random, this is not the case for the *above* and *below* subgroups (see Figure 1). To see this clearly, consider the case of participants "W" and "D", both assigned to the *comparison* group, but while "W" walks to work every day, "D" drives. "W" will be more likely to receive feedback informing her that she is in the *above walking distance* subgroup, while "D" will be more likely to be classified in the *below walking distance* subgroup. Therefore, assignments to these subgroups is not exogenous, but instead individuals self-classify.

The feedback received by individuals in the *comparative* group would gauge them against other "similar" participants, with similarity defined as follows. First, the Euclidean distance between all participants was calculated in a multidimensional space with dimensions sex, age, income, and occupation. Then, the 25% of respondents closer to a participant were labelled as "similar" to him or her. We could not draw similar participants from each individual's social network as our recruiting strategy did not ask participants to encourage their family, friends, and colleagues to join the experiment (as, for example, snowballing procedures do).

The number of calories burned per day presented in the feedback sheet was calculated as a linear function of the distance walked and cycled, considering 85 and 64 Kcal burned per walked and cycled mile, respectively (mapmywalk.com, 2016). The walking energy consumption is consistent with measures by Browning et al. (2006), assuming a 75 Kg person who burns 0.2958 Joules per Kg per walked meter at the preferred speed (1.41 m/s). Such energy consumption rate is the average among men and women of their recommend weight. Calories due to cycling are consistent with the values advised by Ainsworth et al. (2000), assuming vigorous cycling (due to Yorkshire's hilly roads), at 12 mph and a person weighting 75 Kg. Even though these calculations aimed to measure the number of calories burned while travelling, if participants run as an exercise activity, and took their phones with them, then this would be recorded as a new trip, therefore confusing the effect of exercise and travelling. The amount of CO_2 produced by participants was calculated as a linear function of distance travelled by car, bus and rail, considering emissions of 0.13, 0.11, and 0.06 Kg/Km respectively (European Environment Agency, 2016).

2.3 Participants

Participants were recruited in different waves between October 2016 and April 2017, therefore not all respondents provided data for the same weeks. Furthermore, different participants were recruited in different days of the week, and took different amounts of time between completing the first stage, downloading the app, and beginning to use it. Therefore, not all respondents begun their first week of tracking on the same weekday (*i.e.* not everyone's first day was a Monday).

To incentivise participation in the study, respondents who completed the survey in its entirety received a £25 voucher to be used at a major online retailer. Participants were mainly recruited within the greater Leeds area (UK) through mailing lists and flyers.

From an initial sample of 463 individuals who completed the study, 10 were removed from the analysis due to having fewer than two days of tracking after looking at the feedback, or because they were lacking socio-demographic variables. 128 observed days were removed because of errors in the recording of trips, e.g. they reported travel times above 24 hours or average walking speeds above world records.

Our sample is not representative of the general UK population, nor of the Leeds area where it was collected. In particular, our sample has an average income (approx. £52,000 per annum) much higher than the Leeds average (£16,814 pa, Office for National Statistics et al. (2016)).

Women are slightly overrepresented (56% in the sample vs 51% in Leeds, op. cit.), and the amount of one-person households is largely oversampled (56% in the sample as opposed to 33% in Leeds, op. cit.). Table 3 shows the main characteristics of the sample, divided by feedback group. As shown by the p-values of the χ^2 test, also reported in the table, all socio-demographics are independent of the feedback group, meaning that their distribution is equivalent across groups. As our main objective is to test the robustness of results to the effects of endogeneity, we do not require the sample to be representative. However, this also implies that results cannot be readily extrapolated to the population.

2.4 Description of the response variables

Our dataset includes 28,664 trips, of which half are by car and 15% by public transport (see Table 4). This modal split is not representative of the Leeds region (West Yorkshire), where 60% of the commuting trips are performed by car, and 34% by public transport (West Yorkshire Combined Authority, 2016). Table 4 also shows the split of trips by purpose, with going home being the most common purpose, followed by work, leisure/social, and shopping. Students going to University represent only 1% of the recorded trips, reflecting that students are a minority in the sample. Concerning the departure time of these trips, both public transport and car show clear morning and afternoon peaks, while active travel reaches a peak at noon (lunch time, see Figure 3).

Each trip was assigned to a single mode, identified as the one with a longer travel time. However, this was not necessary in the majority of cases, because stops of as little as a minute between legs of the same trip led the *app* to record them as different trips, reducing the issue of mode confusion. For example, walking from home to a train station, waiting there for three minutes and then taking the train to the destination would be recorded as two trips, leading to a very precise measure of distance travelled by each mode. While users could merge trips to indicate they were all part of a single longer trip, this happened rarely, as only 2.6% of trips in the raw data were merged. It was only these trips that needed assignment to the mode used for most of the journey time.

We aggregated the distance travelled by each mode at the day level, providing 2,715 observed days of data. In the data used for our analyses we aggregated the distance and time travelled into three modes: active (walking + cycling), public transport (bus + rail), and car (car + taxi + motorcycle). While information on the distance travelled was available at the mode level (walking, cycling, bus, train and car/taxi/motorcycle), we aggregated it as only few participants cycled (2.4%) and we were not interested in the dynamics of bus and rail substitution, but only in the tendency to use public transport. Additionally, this aggregation allows to clearly rank travel modes according to their social desirability (*i.e.* lack of negative externalities): active travel is preferable to public transport, which is preferable to car.

The distribution of the daily travelled distance by each mode is equivalent on the first and

Table 3: Main characteristics of the sample by feedback group

		Control		Self-monitoring		Comparison		χ^2 test
		(n)	(%)	(n)	(%)	(n)	(%)	p-value*
Female		149	57.1	62	57.4	49	58.3	0.98
Holds university degree		192	73.6	79	73.1	57	67.9	0.58
Full time occupation		167	64.0	71	65.7	49	58.3	0.54
Age	Less than 30	75	28.7	18	16.7	23	27.4	
	30 to 39	77	29.5	37	34.3	19	22.6	
	40 to 49	53	20.3	24	22.2	15	17.9	0.11
	50 to 65	52	19.9	25	23.1	23	27.4	
	More than 65	4	1.5	4	3.7	4	4.8	
Household size	1 person	113	43.3	54	50.0	40	47.6	
	2 people	53	20.3	23	21.3	14	16.7	
	3 people	59	22.6	18	16.7	17	20.2	0.79
	4 people	23	8.8	11	10.2	9	10.7	
	More than 4	13	5.0	2	1.9	4	4.8	
Cars in the household	No car	81	31.0	29	26.9	18	21.4	
	1 car	141	54.0	65	60.2	52	61.9	0.47
	More than 1 car	39	14.9	14	13.0	14	16.7	
Personal income (tens of thousands of £)	Less than 20	51	19.5	15	13.9	20	23.8	
	20 to 40	28	10.7	7	6.5	7	8.3	
	40 to 50	77	29.5	37	34.3	19	22.6	
	50 to 75	53	20.3	24	22.2	15	17.9	0.32
	75 to 100	44	16.9	16	14.8	21	25.0	
	More than 100	8	3.1	9	8.3	2	2.4	
Total		261	100.0	108	100.0	84	100.0	

** p-value of a χ^2 test of independence between sociodemographics and feedback group.*

Table 4: Number and share of trips by mode, purpose and feedback group

	Control	Self-monitoring	Comparison	Total
Active travelling	33.1%	34.2%	36.0%	33.9%
Public transport	15.6%	15.3%	14.7%	15.3%
Car	51.3%	50.5%	49.3%	50.7%
Home	28.4%	28.5%	29.2%	28.5%
Work	19.5%	16.8%	18.2%	18.6%
Leisure/Social	14.3%	13.1%	14.1%	14.0%
Shopping	10.7%	10.2%	11.4%	10.7%
Private business	7.2%	8.8%	7.8%	7.7%
Dropoff/pickUp	7.6%	8.5%	5.5%	7.4%
Change travel mode	6.3%	7.1%	6.4%	6.5%
Exercise	4.1%	5.3%	5.3%	4.6%
Education	1.1%	0.6%	1.2%	1.0%
Petrol	0.6%	0.6%	0.4%	0.6%
Vacation/Travel	0.2%	0.5%	0.5%	0.4%
Total number of trips	16323	7154	5187	28664

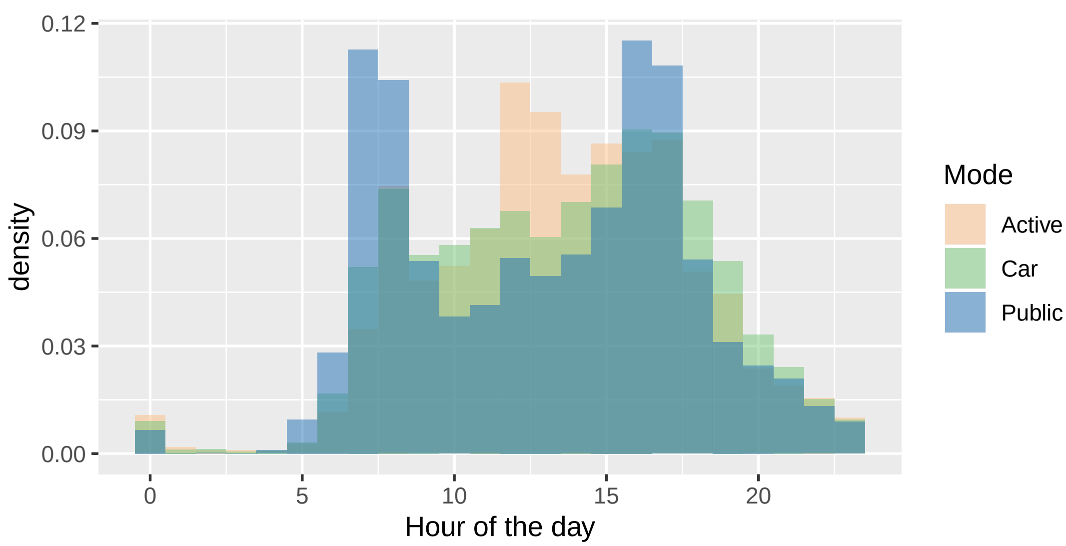


Figure 3: Departure time histograms by mode

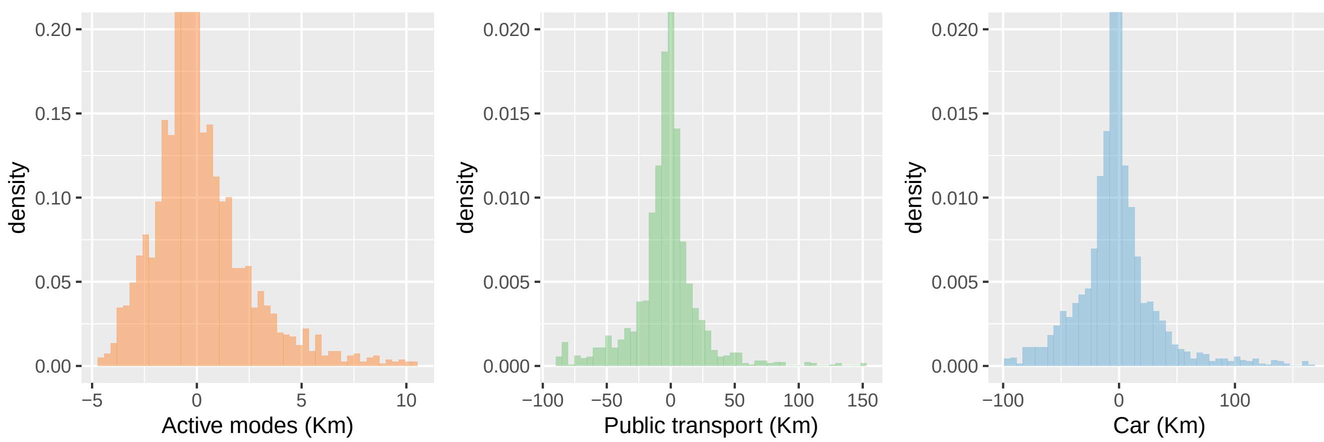


Figure 4: Histograms of the difference in distance travelled *before* and *after feedback*, by mode

second week of tracking, *i.e.* *before* and *after* feedback. We confirmed this using the Kolgoromov-Smirnov test (H0: equivalence), obtaining p-values of 0.35, 0.95, 0.23, and 0.24 for the distance travelled via active modes, public transport, car, and in total, respectively.

Figure 4 shows the distribution of the difference of distance travelled between the *after* and *before* feedback periods, by each mode. More particularly, it depicts the histogram of the difference between the distance travelled each day of the *after* feedback period and the average daily distance travelled during the *before* period, by each mode. For clarity, we leave out the extreme 2.5% of the distribution, and do not show the full height of the peak around zero. All three distributions are skewed to the right, probably as participants increased their trip tagging on the *after* period thanks to the accumulated experience with the tracking application. Quartiles of the distance travelled by each mode in each period, as well as its difference are presented in Table 5. The row "percentage of zeros" in this table represent the amount of individuals who do not use a given mode during a certain period. The amount of zeroes is smaller when calculating differences, as this requires the individual not using the corresponding mode during the whole tracking period. Also, active modes have a much narrower variation range (no more than 60 Km) than public transport and car.

Finally, we collected information about temperature (in Celsius degrees) and rain (share of time raining during the day) for the duration of the experiment, as weather is likely to influence travel behaviour. For simplicity, this data was collected for Leeds only, as 75% of the observed days contain trips in the Leeds area, *i.e.* an approx. 25Km-side square, centred in Leeds city centre.

Table 5: Quartiles of distance travelled by participants (in Km), by mode

	Active			Public Transport			Car		
	Before	After	Δ	Before	After	Δ	Before	After	Δ
Percentage of zeros	40.7%	41.7%	6.2%	67.6%	68.2%	35.3%	32.7%	32.2%	6.2%
1st quartile	1.01	1.01	0.63	8.01	7.63	4.55	9.55	9.14	5.50
2nd quartile	2.21	2.22	1.48	16.17	16.01	9.82	22.67	20.80	14.26
3th quartile	4.42	4.05	3.11	36.19	34.67	24.16	47.25	44.61	35.22
Max	57.52	31.39	30.72	756.93	655.07	647.56	677.41	443.39	427.86

Quantiles are calculated excluding zeros. $\Delta = after - before$

3 Modelling framework

We used a difference-in-difference method (Wooldridge, 2010, Chapter 6) to estimate the effect of information on the distance travelled by different transport modes. The methodology is detailed in this section.

3.1 Generation process

The distance travelled by all individuals by all modes is assumed to share the same data generation process described in eq. (1).

$$y_{int}^{\lambda_i} = \beta_{i0} + \beta_{i1} 1_{t \in A_n} + X_{int}\beta_{iX} + 1_{t \in A_n} \sum_{k=1}^{K} \beta_{ik}\tau_{nk} + c_{in} + \epsilon_{int} \qquad (1)$$

where $y_{int}^{\lambda_i}$ -the dependent variable- is the distance travelled by mode i (active, public transport, or car) by individual n, on day t, under a Box-Cox transformation (see next paragraph). $1_{t \in A_n}$ is a dummy variable equal to 1 if $t \in A_n$, and 0 otherwise; with A_n the set of all days individual n was observed after receiving the feedback. X_{int} is a set of covariates. τ_{nk} is the feedback variable k received by individual n (see next paragraph). c_{in} is an idiosyncratic random error associated to individual n and mode i, whose distribution does not need to be defined. ϵ_{int} is an independent identically distributed (iid) normal random error with mean zero and standard deviation σ_i. All β parameters must be estimated. β_{i0} is an intercept and β_{i1} is the average change between the *before* and *after* feedback periods, capturing any systematic differences between the two periods, such as increased familiarity with the tracking app. β_{iX} captures the effect of covariates and β_{ik} the effect of feedback variables.

There is substantial heterogeneity in the daily distance travelled by participants (see Table 5). Variability increases the difficulty of measuring the effect of feedback, as the scale of changes can be very different between individuals. For example, consider participants "N" and "F" who commute by car. While "N" lives near his/her workplace, "F" lives far from it. If both change their commuting mode to public transport, the impact on the total distance travelled by each mode will be small for "N" but large for "F". We take this phenomenon into consideration by applying a two-parameter Box-Cox transformation (Box and Cox, 1964) to the distance travelled by each mode. The transformation is as follows.

$$y_{int}^{\lambda} = \begin{cases} \frac{(y_{int}+\lambda_{i2})^{\lambda_{i1}}-1}{\lambda_{i1}} & \lambda_{i1} \neq 0 \\ \log(y_{int}+\lambda_{i2}) & \lambda_{i1} = 0 \end{cases} \tag{2}$$

where y_{int} is the distance travelled by mode i by individual n, on day t, and $\lambda_i 1$ and $\lambda_i 2$ are parameters to be estimated. These parameters were estimated by maximizing the likelihood of eq. 1. While we estimate different transformations for each mode, we do not make any difference between individuals (n) or days (t). Henceforth, every time we refer to the dependent variables, we are referring to its Box-Cox transformation.

Concerning the feedback variables (τ_{nk}), we consider four of them ($K = 4$).

- $\tau_{n,self}$: Takes the value 1 if individual n belongs to the *self-monitoring* group, *i.e.* received feedback only about his/her own travel behaviour; it is equal to 0 in any other case. We used a dichotomous variable instead of the value of the different variables that were displayed as part of the feedback as we do not know what the reference point used by the individual is. In other words, we do not know if the participant considers his/her distance and time spent travelling as high or low.

- $\tau_{n,distWalkAbove}$: For participants in the *comparison* group who walked a longer distance than others like them, this variable takes the value of the additional distance travelled (in Km) with respect to other similar participants under a transformation. It takes a value of 0 in any other case. More formally:

$$\tau_{n,distWalkAbove} = \begin{cases} \log(1 + \max(\text{self}_{n,distWalk} - \text{others}_{n,distWalk}, 0)) & n \in \textit{comparison} \text{ group} \\ 0 & \text{in other case} \end{cases}$$

  where $\text{self}_{n,distWalk}$ is the average daily distance walked by participant n during her *before feedback* period, while $\text{others}_{n,distWalk}$ is the average daily distance walked by similar participants during their *before feedback* periods. We tested other transformations of the value (e.g. linear, quadratic), but the log transformation provided better fit.

  For example, if participant n from the *comparison* group walked a daily average of 8 Km during her *before feedback* period (*i.e.* first week of tracking), while similar participants walked 5 Km in their first week, then $\tau_{n,distWalkAbove} = \log(1+3)$. On the other hand, if she only walked 3 Km on average while others walked 5, then $\tau_{n,distWalkAbove} = \log(1+0) = 0$. This variable always takes the value 0 for individuals in the *control* and *self-monitoring* groups.

- $\tau_{n,distCarAbove}$: Analogous to $\tau_{n,distWalkAbove}$, but concerning the distance (in Km) travelled by car. This variable takes a value of zero if *car* is unavailable for respondent n's household.

- τ_{n,CO_2Above}: Analogous to $\tau_{n,distWalkAbove}$, but concerning the CO_2 emission (in Kg) due to travelling, and without using a logarithmic transformation. More formally:

$$\tau_{n,CO_2Above} = \begin{cases} \max(\text{self}_{n,CO_2} - \text{others}_{n,CO_2}, 0) & n \in \textit{comparison} \text{ group} \\ 0 & \text{in other case} \end{cases}$$

Nine other feedback variables were tested in our modelling, but are not described above. Some of them were analogous to the ones described, but dealt with how many kilometres and hours participants walked or drove, above or below the average of others. Other similar variables measured how many calories and Kg of CO_2 participants burned or emitted above or below the average of others. Several of these feedback variables were highly correlated with the ones we describe above (e.g. distance and time walked had a correlation of 0.85), leading to very high Variance Inflation Factor (O'Brien, 2007) if both were included in a model. The feedback variables described above are the ones that performed better in modelling, both in terms of significance, fit and VIF.

We did not consider additional feedback variables related to cycling, bus and train riding, due to the trip share of these modes being significantly lower than that of walking and travelling by car (see Table 4). Likewise, we did not consider any feedback variable relating to the number of out-of-home activities performed by participants. As the need for travel is derived from the activities performed by an individual, explaining the amount of travelling based on the number of out-of-home activities performed would be highly endogenous.

As the demand for transport is derived from the activities an individual performs, we expect substitution and complementarity between the distances travelled by each mode. This is because changing travel modes is in many cases easier than changing activity patterns. For example, a participant can change her commuting mode from bus to cycling relatively easily, but she can hardly stop going to work. In terms of modelling, this implies that the error term ϵ_{int} in eq. (1) is most likely correlated among different responses $y_{int}^{\lambda_i}$. We take this into consideration by assuming the vector $[\epsilon_{active,nt}, \epsilon_{public,nt}, \epsilon_{car,nt}]$ follows a multivariate normal distribution with mean zero and a variance-covariance matrix Σ to be estimated.

3.2 Difference-in-difference

If the generation process of the data is the same in both periods, then we can take the difference between them and still consistently estimate the effect of the feedback β_{ik} (Wooldridge, 2010, Chapter 6). Focusing on the difference between periods implies dropping all terms that are constant across periods, *i.e.* both the observable (X_{int}) and unobservable (c_{in}) characteristics of the individuals. This is beneficial as it implies we do not need to control for any covariate (observable or otherwise) that remains constant across periods, as long as they influence the response in a linear fashion and do not interact with the feedback variables. However, this also means that the number of observations is reduced to the number of *after* - *before* pairs we can construct from our data.

There are several ways in which the *after* - *before* pairs can be constructed. One possibility would be to match each *after feedback* day with its corresponding day in the *before feedback* period, e.g. matching Monday with Monday, Tuesday with Tuesday, etc. However, there is little evidence in our data of consistent travel patterns between corresponding days of the week once we control for individual effects: distances travelled by each mode have correlations smaller than 0.06 between corresponding days. More importantly, the feedback provided to participants summarised the aggregate travel patterns of the whole *before feedback* period, not of each individual day during that period. Therefore, it is more likely that during the *after feedback* period individuals compare their daily travel behaviour against the average of the *before feedback* period, rather than to their behaviour on a single day of it. Consequently, we pair each day of the *after feedback* period with the *average* of the *before feedback* period. This approach also avoids dropping any observation from the *after* period, as it will always be possible to pair it with the *average*. More formally, $\forall t \in A_n$:

$$\begin{aligned}
\Delta y_{int}^{\lambda_i} &= y_{int}^{\lambda_i} - \frac{1}{d_n} \sum_{t' \in B_n} y_{int'}^{\lambda_i} \\
&= \beta_{i1} + \Delta X_{int} \beta_{iX} + \sum_{k=1}^{K} \beta_{ik} \tau_{nk} + \epsilon_{int} - \frac{1}{d_n} \sum_{t' \in B_n} \epsilon_{int'} \\
&= \beta_{i1} + \Delta X_{int} \beta_{iX} + \sum_{k=1}^{K} \beta_{ik} \tau_{nk} + \epsilon_{int} - \tilde{\epsilon}_{in} \qquad (3)
\end{aligned}$$

where B_n is the set of days t in the *before feedback* period, and d_n its cardinality. The idiosyncratic component c_{in} is absent from (3) as it does not change from one period to the next. As before, ϵ_{int} follows a $N(0, \Sigma)$ distribution, and is independent across observations. On the other hand, $\tilde{\epsilon}_{in}$ follows a $N(0, \frac{1}{d_n}\Sigma)$ distribution, and is common across observations of the same individual.

In eq. 3, only the covariates that change from one period to the next are relevant in ΔX_{int}, as all others become zero. This implies, for example, that all socio-demographic characteristics of the participants drop from eq. 3, as we can safely assume they do not change between periods. We include only three explanatory variables in ΔX_{int}. The first one ($weekend_{nt}$) is a dummy variable taking the value 1 if day t for participant n is a Saturday or Sunday, and takes the value 0 otherwise. We included this variable as previous studies have found significantly different travel patterns on weekends as compared to the rest of the week (Wang et al., 2018). The second explanatory variable ($\Delta Temp_{nt}$) is a continuous variable indicating the difference in temperature (in Celsius degrees) between day $t \in A_n$, and the average temperature of the *before feedback* period. The third explanatory variable ($\Delta rain_{nt}$) is a continuous variable between -1 and 1, indicating the difference between the share of time during day t when it was raining, and the average of the same share during the *before feedback* period.

Working in differences imply assuming the "parallel trends" hypothesis, *i.e.* that the generation process (eq. 3) is the same in the *before* and *after* period, and between all feedback groups (*control*, *self-monitoring*, and *comparison*). This is reasonable as we do not expect participants' contexts to change significantly from one week to the next, neither across groups, as they have equivalent socio-demographic characteristics. Increasing familiarity with the tracking app is not relevant, as we only included participants who provided details for at least 95% of their trips over both periods, and any further familiarity effect would be captured by the constant in eq. 3.

3.3 Endogeneity correction

For τ_{nk} to be correctly estimated, the assignment to a feedback group must be exogenous, a condition achieved if the assignment is random. While assignment to the *control*, *self-monitoring*, and *comparison* groups is random, assignment to *comparison*'s subgroups is not. Therefore, the feedback variables $\tau_{n,distWalkAbove}$, τ_{n,CO_2Above} and $\tau_{n,distCarAbove}$ are endogenous. In these cases, the feedback received is likely to correlate with the error term, causing an endogeneity problem. To see this more clearly, consider the case of participants "S" and "A", both of which are assigned to receive *comparative* feedback, but while "S" lives a sedentary life, "A" is a very active cyclist. As "S" is sedentary, he will likely drive more than others, and therefore have $\tau_{S,distCarAbove} > 0$, while instead, "A" will more likely have $\tau_{A,distCarAbove} = 0$. At the same time, because "S" is sedentary, he will be less likely to reduce his driving due to the feedback, while "A" might be much more encouraged to reduce her driving if she was told she drives for longer than others. In other words, $\Delta y_{int}^{\lambda_i}$ will depend on unobservable attributes (e.g. lifestyle) that also correlate with explanatory variable ($\tau_{n,distCarAbove}$), generating endogeneity issues.

The issue of endogeneity could be avoided by providing randomized information to participants in the *comparison* group, *i.e.* contrasting participants' behaviour against that of an artificially and randomly constructed peer. While this would make the feedback variables exogenous, it would involve presenting false information to participants, thus raising ethical questions.

We correct for endogeneity using the two-stages least square method (2SLS), as described by Wooldridge (2010, Chapter 21). This methods implies replacing the value of the endogenous variables by linear projections of them onto an exogenous space. In other words, it requires the estimation of one additional linear model for each endogenous variable, as shown in eq. (4).

$$\tau_{nk} = \Delta X_{int}\alpha_{k,X} + Z_{ink}\alpha_{k,Z} + \nu_{nk} \tag{4}$$

$$\Delta y_{int}^{\lambda_i} = \beta_{i1} + \Delta X_{int}\beta_{iX} + \sum_{k=1}^{K}\beta_{ik}\hat{\tau}_{nk} + \epsilon_{int} - \tilde{\epsilon}_{in} \tag{5}$$

where Z_{ink} is a set of instruments for τ_{nk}, *i.e.* a set of explanatory variables that correlate with τ_{nk} but do not correlate with $\tilde{\epsilon}_{in}$ (in our case, socio-demographic and attitudinal variables). α_{kZ} and α_{kX} are vectors of parameters to be estimated, and ν_{nk} is an iid normal random error with mean zero and a standard deviation of σ_k, also to be estimated. $\hat{\tau}_{nk}$ is the prediction of τ_{nk} according to eq. 4, *i.e.* $\hat{\tau}_{nk} = \Delta X_{int}\alpha_{k,X} + Z_{ink}\alpha_{k,Z}$. The endogeneity correction is only relevant for individuals who receive endogenous feedback; for example eq. (4) applies to $\tau_{n,walkDistanceAbove}$ only for those individuals in the *comparison* group who walked further than others. Equation (4) never applies to individuals in the *control* and *self-monitoring* groups. In other words, eq. (4) applies only when $\tau_{nk} > 0$. Equation (5) is analogous to eq. (3), but replacing τ_{nk} by $\hat{\tau}_{nk}$; eq. (5) is used when the endogeneity correction is needed, and eq. (3) when it is not.

We include participants' unobserved attitudes towards social influence among the instruments in Z_{ink}. We measure them using a Structural Equation Model (SEM), as discussed by Bollen (1989). We use a linear structural equation (eq. 6) for the attitude level, with individuals' socio-demographics (X_n) explaining the level of the attitudes. We also use three indicators for each attitude (*i.e.* latent variable), where each indicator is the level of agreement with a statement. We use linear links between the indicators and the attitude level (eq. 7).

$$a_{ln} = X_n\gamma_l + \eta_{ln} \tag{6}$$

$$I_{lmn} = \lambda_{lm}a_{ln} + \xi_{lmn} \tag{7}$$

where l enumerates attitudes and m indicators. a_{ln} is the measurement of attitude l for individual n, while I_{lmn} is individual n's level of agreement with indicator m of attitude l. η_{ln} and ξ_{lmn} are random error components, both following normal distribution with mean zero, the first one with a fixed standard deviation of 1, and the second one with a standard deviation of σ_{lm} to be estimated. γ_l and λ_{lm} are parameters to be estimated.

4 Results

The results section is organized into four subsections, with the first one presenting the parameters of the Box-Cox transformation applied to the dependent variables, and the following estimating the effect of feedback, each in an increasingly detailed way. To estimate the feedback effect, we first use eq. (3) directly, assuming ϵ_{int} to be uncorrelated across different modes. Then, we again use eq. (3), but this time assuming correlation between different modes, as described in section 3.2. Finally, we estimate the effect of feedback considering correlation between modes and correcting for endogeneity.

All models were estimated using Simulated Maximum Likelihood, using 500 Halton draws to simulate the $\tilde{\epsilon}_{in}$ and η_n random components. Estimation was performed in R (R Core Team, 2018), using the package Apollo (Hess and Palma, 2019). When considering endogeneity correction, estimation of all models (equations 4, 5, 6 and 7) was performed simultaneously to ensure consistent standard errors.

Measuring the fit of models with random components is not as straightforward as with traditional linear regressions. The main issue is that the traditional R^2 does not take into consideration the effect of random components. To solve this shortcoming, we use the fit measurements proposed by Nakagawa and Schielzeth (2012). We use two measures of fit: R^2 *GLMM (m)* that is equivalent to the traditional R^2, and R^2 *GLMM (c)* which takes the randomness into account. The idea behind these measurements is to express the percentage of variance explained by the model based on the estimated standard deviations, as opposed to the data and model predictions. Still, both of these fit measures ignore the effect of correlation between dependent variables. Therefore, we still consider the model log-likelihood as the most trustworthy indicator of fit.

4.1 Box-Cox transformation

As mentioned in section 3.1, all dependent variables underwent a Box-Cox transformation (eq. 2). Parameters of these transformations are presented in Table 6.

Table 6: Box-Cox transformation parameters for dependent variables

Distance by mode	λ_1	λ_2
Active travelling	0.0000	0.0006
Public transport	0.0000	0.0076
Car	0.1310	0.0068

$\lambda_1 = 0$ *implies a* log *transformation*

4.2 The effect of feedback: ignoring correlation and endogeneity

The most straightforward approach to estimate the effect of feedback on the travel behaviour of participants is by independently estimating equation (3) for the distance travelled by each mode. Results from this approach are presented in Table 7. According to this table, only three feedback variables (τ_{nk}) have a significant effect: (i) individuals receiving information about their own travel behaviour reduce their distance travelled by active modes (*i.e.* walking and cycling), except for female participants; (ii) individuals who were told that they used active modes for longer distances than others reduce their use of them; and (iii) individuals told they drive longer distances than others reduce their driving. Previous studies would have seen these results as confirming that information can alter behaviour.

The only significant heterogeneity in the effect of feedback information was found among the *self-monitoring* group. We found that while most participants tend to reduce their active travelling distance after receiving information about their own travel behaviour, car owners do it to a lesser extent, and female participants actually exhibit the opposite effect. The effect on car owners is reasonable, as it may be easier for them to replace short car trips by walking or cycling. Individuals who do not own a car, on the other hand, are more likely to already walk short distances, to avoid the cost and waiting time of public transport. On the other hand, we cannot explain the effect on female participants with the available information. We tested other forms of heterogeneity, namely interactions of the feedback variables with additional socio-demographics and attitudes, as well as random coefficients. We did not find any significant variance on the random coefficients (assumed to follow a normal distribution), and neither did we find any other significant interaction.

The controls included in the regressions indicate that participants tend to use active and public travel modes more intensely on weekdays rather than weekends, while the opposite is true for car. This is reasonable as the high flexibility of the car makes it attractive for weekend leisure trips. Furthermore, weekend trips tend to involve multiple household members, reducing its per capita cost. No other control is significant, though rain comes close with a t-test of -1.54 (p-value = 0.12), and a negative effect on the active travel modes. The same controls were included for all modes, despite their insignificance, to ease comparison. Fit indices are compared across models in the discussion section.

The low fit in our models is not a major concern. We are not interested in predicting behaviour, but only in measuring the effect of our intervention, and to that end, we simply want to ensure the consistency and significance of our estimates. This is equivalent to the use of ANOVA testing in other studies (Jariyasunant et al., 2015; Weber et al., 2018), where global fit of the model is usually not even reported. The significance of parameters, on the other hand, can be correctly assessed in our models by their corresponding t-ratios. Furthermore, we are reassured that our explanatory variables do influence behaviour by the fact that all models fit significantly better than models with constants only, according to a Likelihood ratio test.

Table 7: Parameter estimates and fit indices ignoring correlation and endogeneity

	Δ Active distance		Δ P.T. distance		Δ Car distance	
	Coeff.	(t-ratio*)	Coeff.	(t-ratio*)	Coeff.	(t-ratio*)
Intercept	-0.0711	-0.58	0.0761	0.72	0.1818	1.47
Weekend	-0.7248	-4.08	-1.2753	-7.84	0.6557	3.55
Δ Temperature	-0.0175	-0.69	-0.0231	-1.08	0.0067	0.25
Δ Rain	-0.4716	-1.54	0.0553	0.23	0.1716	0.62
Self-monitoring	-0.8563	-2.42	0.2542	1.48	-0.3778	-1.79
x owns car	0.6494	2.02				
x female	1.0397	3.15				
walk dist (above) †	-1.3473	-3.52				
CO_2 (above)			-0.0575	-1.24		
car dist (above) †					-0.2499	-2.91
σ	3.2965	66.55	2.7516	47.38	2.6740	38.17
R^2 GLMM(m) ‡	0.023		0.039		0.018	
R^2 GLMM(c) ‡	0.145		0.159		0.141	
LL	-7222.17		-6731.31		-4798.53	
LL full model						-18752.00
LR test against constants only (p-value)					206.07	(0.00)
AIC						37556.69
BIC						37692.54

** Robust t-ratios. † Transformed as log(1+x) ‡R^2 ignores correlation.*

4.3 The effect of feedback: considering correlation and ignoring endogeneity

Considering correlation between the error terms in eq. (3) induces little change into the estimates, shown in Table 8. As before, only two pieces of feedback have a significant effect: telling participants they walked more than others, and driven more than others, leading to a reduction in active travel and travelling by car, respectively. The effect of controls remain the same, with active and public transport modes being used less over the weekends.

The correlation pattern between error terms provides the expected results: active travel and public transport are complementary modes, with their error terms having a correlation of 0.0987. On the other hand, car is a substitute for both active travel and public transport, showing a correlation of -0.1618 and -0.3413 with each, respectively. The reason these values are relatively low is because we are working in differences. We repeated the analysis using eq. (1) and found a stronger correlation pattern: 0.05 between active and public modes, -0.23 between active and car, and -0.48 between public and car.

Allowing for complementarity and substitution in the model leads to an improvement in fit of 152 log-likelihood points. This difference is significant ($p<0.01$) according to a Likelihood-ratio test. However, the R^2 indices worsen as these measures ignore the effect of correlation between the dependent variables.

4.4 The effect of feedback: considering correlation and endogeneity

As discussed in section 3.3, correcting for endogeneity requires estimating equations (4), (5), (6) and (7) simultaneously.

Figure 5 presents the structure of the SEM model, *i.e.* of equations (6) and (7). There are three unobservable attitudes (*i.e.* latent variables): *impressions*, *approval* and *emulation*, each relating to a different social need. The first attitudes relates to the need for making a good impressions on others. The second relates to the need for the approval of others. And the third relates to the tendency of emulating the behaviour of others. Each attitude is explained by participants' characteristics, and in turn explains the level of agreement that they manifested with a series of statements.

Table 9 presents the parameters of the SEM model. All explanatory variables are highly significant. Older individuals tend to score higher on *impressions*. Female participants scored higher on *approval* and *emulation*. Individuals with a university degree score higher on all three attitudes. Full time workers score higher in *approval*. Finally, having a higher income (measured thousand of pounds per year) also increase the score of *approval*. As our objective is using these attitudes as instruments in eq. (4), and not to study their effect on travel behaviour in particular, nor study their determinants, we do not discuss them further.

Table 8: Parameter estimates and fit indices considering correlation and ignoring endogeneity

	Δ Active distance		Δ P.T. distance		Δ Car distance	
	Coeff.	(t-ratio*)	Coeff.	(t-ratio*)	Coeff.	(t-ratio*)
Intercept	-0.0706	-0.57	0.0776	0.74	0.182	1.54
Weekend	-0.7260	-4.09	-1.2757	-7.84	0.6061	3.42
Δ Temperature	-0.0167	-0.66	-0.0233	-1.08	-0.0014	-0.05
Δ Rain	-0.4685	-1.52	0.0501	0.21	0.0432	0.16
Self-monitoring	-0.8501	-2.34	0.2486	1.46	-0.3842	-1.85
x owns car	0.6747	2.02				
x female	0.9908	2.99				
walk dist (above) †	-1.3358	-3.36				
CO_2 (above)			-0.0626	-1.47		
car dist (above) †					-0.1931	-2.28
Sigma	3.2956	66.45	2.7504	47.83	2.6693	46.55
Correl with Δ Act.			0.0987	4.09	-0.1618	-5.27
Correl with Δ Pub.	0.0987	4.09			-0.3413	-9.72
Correl with Δ Car	-0.1618	-5.27	-0.3413	-9.72		
R^2 GLMM(m) ‡	0.022		0.040		0.015	
R^2 GLMM(c) ‡	0.144		0.160		0.138	
LL						-18599.51
LR test against constants only (p-value)					167.00	(0.00)
AIC						37251.02
BIC						37404.59

** Robust t-ratios. † Transformed as log(1+x) ‡R^2 ignores correlation.*

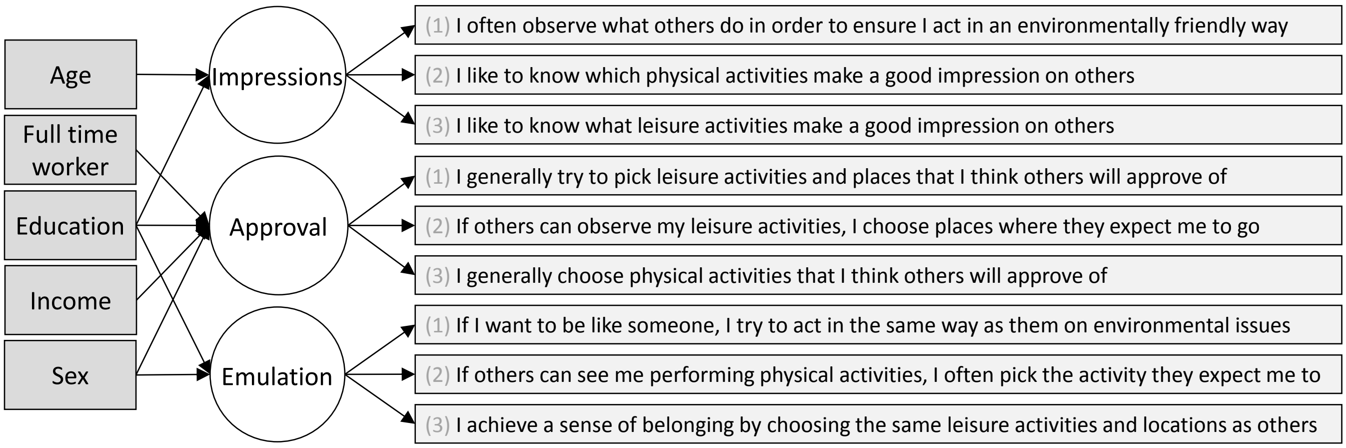


Figure 5: Structure of the SEM model

Table 9: Parameter estimates and fit indices of the SEM model

	Impressions		Approval		Emulation	
	Coeff.	(t-ratio*)	Coeff.	(t-ratio*)	Coeff.	(t-ratio*)
Age	0.0767	3.42				
Female			0.5958	9.1	1.1714	34.22
University degree	1.3867	4.24	1.2791	4.96	1.1268	21.14
Works full time			0.6879	16.2		
Personal income			0.0188	16.17		
λ_1	1.1554	3.91	1.6598	19.45	2.5171	31.04
λ_2	0.9854	3.88	1.4707	19.03	1.8609	30.08
λ_3	0.9225	3.9	1.4677	19.01	1.9069	29.57
σ_1	1.0172	17.01	0.806	16.73	1.0189	18.72
σ_2	1.1126	16.69	0.8802	19.79	1.055	23.7
σ_3	1.0283	23.23	0.7906	16.53	0.9973	23.67
LL	-2212.26		-2067.11		-2468.90	

** Robust t-ratios. †R^2 ignores correlation.*

Estimation results for eq. (4) are presented in Table 10. The objective of this equation is to remove the endogenous part of the feedback variables by projecting them into an exogenous space of instruments, *i.e.* by explaining them based on exogenous variables. We found appropriate instruments for all endogenous feedback variables: *emulation* for $\tau_{i,walkDistAbove}$; full time worker, number of cars, *impressions*, and *approval* for τ_{i,CO_2Above}; and level of education, *approval* and *emulation* for $\tau_{i,carDistAbove}$. As required by the 2SLS methodology, all covariates included in eq. (5) are also included as explanatory variables, despite their insignificance.

We are not interested in the effect of the instruments on the endogeneous variable, but only on their capacity to explain it to a high enough degree. All three endogenous variables achieve appropriate fit, with *car distance (above)* reaching the lowest: an R^2 GLMM (m) of 0.275, and an R^2 GLMM (c) of 0.365.

Table 10: Parameter estimates and fit indices of the first stage of 2SLS

	walk dist (above) †		CO_2 (above)		car dist (above) †	
	Coeff.	(t-ratio*)	Coeff.	(t-ratio*)	Coeff.	(t-ratio*)
Intercept	-1.0083	-22.19	12.6897	12.57	5.2039	328.29
Weekend	-0.0034	-0.32	0.1344	0.88	-0.0181	-0.98
Δ Temperature	-0.0019	-0.51	0.0397	0.59	-0.0136	-1.23
Δ Rain	-0.0205	-0.72	-0.3498	-1.15	0.0025	0.05
Age			-0.1042	-2.27		
University degree					1.1498	3.71
Impressions (latent)			2.0582	3.43		
Approval (latent)			-3.7166	-7.3	-1.2824	-143.7
Emulation (latent)	0.9031	75.58			0.4382	14.41
Sigma	0.0613	7.26	0.6838	6.55	0.1181	6.53
R^2 GLMM(m) ‡	0.352		0.247		0.203	
R^2 GLMM(c) ‡	0.433		0.341		0.303	
LL	95.08		-184.30		72.84	

** Robust t-ratios. † Transformed as log(1+x) ‡R^2 ignores correlation.*

Finally, Table 11 presents the parameter estimates and fit indices of the regression measuring the impact of feedback information on travel behaviour when considering both correlation and the endogeneity correction. In this table, we see that no feedback variable has a significant effect on participants' travel behaviour. In particular, the relevant t-ratios suffer a steep decrease from the case without endogeneity correction, shrinking by 3.32, 1.41 and 0.56 points towards zero for active, public, and car modes, respectively. The magnitude of the feedback coefficients also shrink towards zero. The effect of controls as well as the correlation pattern between dependent variables remain largely the same to the one in previous models.

The model correcting for endogeneity has a worse fit (R^2, log-likelihood, AIC and BIC) than the model considering only correlation between alternatives (see Table 8). This is due to the 2SLS method replacing explanatory variables with exogenous projections of them (through the use of eq. 4). This increases the noise in the model, decreasing fit. But beyond the lower fit, the model with endogeneity correction increases our certainty of parameters being consistent.

Table 11: Parameter estimates and fit indices considering correlation and endogeneity

	Δ Active distance		Δ P.T. distance		Δ Car distance	
	Coeff.	(t-ratio*)	Coeff.	(t-ratio*)	Coeff.	(t-ratio*)
Intercept	-0.1711	-0.71	0.0438	0.15	0.7441	1.97
Weekend	-0.7273	-4.09	-1.2842	-7.88	0.5791	3.25
Δ Temperature	-0.0103	-0.39	-0.0286	-1.29	-0.0034	-0.13
Δ Rain	-0.6183	-1.95	-0.0448	-0.18	0.0179	0.06
Self-monitoring	-0.9971	-2.49	0.2785	1.5	-0.2135	-0.97
x owns car	0.8017	2.11				
x female	1.0312	2.7				
walk dist (above) †	-0.0108	-0.04				
CO2 (above)			-0.0025	-0.06		
car dist (above) †					-0.2066	-1.72
Sigma	3.3514	66.55	2.7928	47.86	2.7153	46.72
Correl with Δ Act.			0.1035	4.16	-0.1640	-5.34
Correl with Δ Pub.	0.1035	4.16			-0.3462	-9.67
Correl with Δ Car	-0.1640	-5.34	-0.3462	-9.67		
R^2 GLMM(m) ‡	0.016		0.037		0.012	
R^2 GLMM(c) ‡	0.139		0.157		0.135	
LL						-18636.92
LR test against constants only (p-value)					92.18	(0.00)
AIC						37325.84
BIC						37479.45

** Robust t-ratios. † Transformed as log(1+x) ‡R^2 ignores correlation.*

5 Discussion

The objective of this research was to examine the role of information provision on travel behaviour. In particular, we examined two persuasion strategies: providing *self-monitoring* and *comparison* information through a smartphone app. The first strategy consisted in providing individuals with information about their own travel behaviour. The second strategy was to provide a comparison with similar individuals in addition.

Our experimental design did not include any gamification strategy, such as proposing goals and rewards to participants (Di Dio et al., 2018; Piras et al., 2018), inducing direct competition between them through leader boards (Barrat, 2017), or fostering collaboration through team quests or achievements (Weber et al., 2018). Neither did we provide specific suggestions to change behaviour (Bucher et al., 2019; Meloni et al., 2013). Instead, we passively provided information to participants in a single occasion, in a similar fashion to Jariyasunant et al. (2015).

Our results consistently indicate that the provision of *self-monitoring* information induces a decrease in the use of active travel modes among men, and does not have an effect on women. This results is the same no matter the modelling approach, though its significance lowers as the modelling becomes more complex. The effect on men could be explained as a case of *moral licensing* (Khan and Dhar, 2006), where the individual believes he walks or cycles more than enough, and therefore reduces his active travel. On the other hand, the case of women could point to the information being rendered useless due to the lack of context.

It is difficult to directly compare our results with other studies evaluating the efficacy of the *self-monitoring* strategy. Most other studies involve other gamification strategies, notably *praise* and *competition* (Barrat, 2017; Di Dio et al., 2018; Weber et al., 2018), and also a constant provision of information, as opposed to an isolated intervention as in our case. The three studies just cited report relevant changes in the travel behaviour of participants, though Barrat (2017) only performs a qualitative analysis, Di Dio et al. (2018) work with a small (77) sample of post-graduate students whose commuting is less than 10 Km long, and Barrat (2017) and Weber et al. (2018) work with a highly involved sample of cyclists. These results hint to the necessity of continuous information provision, and high levels of involvement to induce a change in travel behaviour. Gamification strategies could be a useful way to achieve both but, to the best of our knowledge, they have only been tested on strongly biased population.

The effect of providing *comparison* information change depending on the modelling approach. If endogeneity is ignored, then the *comparison* information has two significant effects: participants who are told that they walk (or cycle) more than others tend to reduce their walking (or cycling) in the near future. Also, participants informed that they drive further than others decrease their driving in the near future. While the first result is not reported by any other study, it could be an example of moral licensing (Khan and Dhar, 2006). On the other hand, the reduction in driving does match the results of Jariyasunant et al. (2015), who finds a reduction of 33% in their sample of 78 individuals receiving *comparison* feedback. As they do not differentiate between participants above and below the behaviour of others, a more detailed comparison is not possible.

However, when correction for endogeneity is implemented, all effects of the *comparison* information provision become insignificant. This points to the results discussed in the previous paragraph being an artefact of model estimation, and not a real consequence of the information provision. The issue of endogeneity has been addressed in studies dealing with social influence (Maness and Cirillo, 2016; Walker et al., 2011), but it is usually ignored when measuring the effect of information provision. When evaluating interventions in the form of information provision, most studies report that increased engagement with the corresponding tool is a good predictor of change in behaviour (Barrat, 2017; Jariyasunant et al., 2015; Weber et al., 2018). This could be an indicator of the presence of endogeneity. Participants more engaged with the tool are probably those more interested in reducing their environmental impact in the first place. Therefore, the effects could be due to individuals' intrinsic (*i.e.* self) motivation to reduce their car use and increase their active travelling, and not a product of the information provision.

If so, then a more promising approach for tools promoting active travel (e.g. smartphone apps) could be fostering extrinsic motivation among individuals who otherwise would not engage with active travel. For example, Di Dio et al. (2018) propose an *app* that awards points for their use of active travel modes, which can later be spent on local retail stores. The effect of these points on behaviour is less likely to be endogenous, as the desire to acquire the points is driven by their usefulness more than the individual's intrinsic motivation to use active travel modes.

Finally, it is important to recognise the limitations of this study. First, the sample size is relatively small, especially for the *comparison* feedback group. Secondly, it is conceivable that a single round of feedback is not sufficient to induce behavioural change, and also that a single week of tracking after providing the information is not enough for participants to significantly change their behaviour (e.g. plans might have already been laid for the second week). However, in the context of a two-week survey, multiple rounds were not judged to be feasible. Longer survey periods are of course possible, but would likely increase sample attrition and require analysts to reduce the level of detail in the survey (Axhausen et al., 2002; Schlich et al., 2002). A more longitudinal non-contiguous approach may also be useful, where tracking could be performed for several weeks, but allowing for long intervals without tracking between them. Thirdly, the delay in looking at the feedback by some participants may have reduced its impact, as they may have considered it no longer relevant. Finally, a more engaging design of the information delivery in our experimental design may have increased the effect of the information. Continuous provision of information and additional gamification could contribute to larger effects. Still, even under those conditions, a thorough analysis including endogeneity correction would still be needed to assess the impact of such information provision strategies. In the same vein, providing information about "other people like yourself" may not be as enticing as providing information about each participant's own social network. The same is true for *simulation* persuasion strategies, where consequences of decisions are communicating before making a choice. Indeed, Avineri and Goodwin (2010) claims that "if individuals are unable to equate current actions with consequences, then changes may be less significant", which could explain our findings. Hence, our negative results should not necessarily be extended to situations with more arresting information delivery strategies.

In conclusion, simply providing isolated information about individuals' own travel behaviour, or about themselves as compared to others, is not enough to increase active travel nor diminish car use. Interventions with a higher level of engagement, especially those promoting extrinsic motivation, could be more promising, such as gamified information systems including rewards and providing environmentally friendly travel alternatives in real time. Still, a rigorous measurement of efficacy is required in those conditions, through a long period of time, and on a sample representative of the population. The core message of our work is however that even in these circumstances, analysts should be mindful of mis-inferring effects and should use the endogeneity corrections discussed in our paper.

6 Acknowledgements

The Leeds authors acknowledge the financial support by the European Research Council through the consolidator Grant 615596-DECISIONS.

References

Barbara E Ainsworth, William L Haskell, Melicia C Whitt, Melinda L Irwin, Ann M Swartz, Scott J Strath, WILLIAM L O Brien, David R Bassett, Kathryn H Schmitz, Patricia O Emplaincourt, et al. Compendium of physical activities: an update of activity codes and met intensities. *Medicine and science in sports and exercise*, 32(9; SUPP/1):S498–S504, 2000.

Evangelia Anagnostopoulou, Babis Magoutas, Efthimios Bothos, Johann Schrammel, Rita Orji, and Gregoris Mentzas. Exploring the links between persuasion, personality and mobility types in personalized mobility applications. In Peter W. de Vries, Harri Oinas-Kukkonen, Liseth Siemons, Nienke Beerlage-de Jong, and Lisette van Gemert-Pijnen, editors, *Persuasive Technology: Development and implementation of personalizes technologies to change attitudes and behaviors*, pages 107–118. 12th International conference PERSUASIVE 2017, 2017. doi: 10.1007/978-3-319-55134-0.

Erel Avineri and Phil Goodwin. Individual behaviour change: evidence in transport and public health. techreport, The Department for Transport, London, UK, 2010.

Kay W. Axhausen, Andrea Zimmermann, Stefan Schönfelder, Guido Rindsfüser, and Thomas Haupt. Observing the rhythms of daily life: A six-week travel diary. *Transportation*, 29(2): 95–124, 2002. doi: 10.1023/a:1014247822322.

Paul Barrat. Healthy competition: A qualitative study investigating persuasive technologies and the gamification of cycling. *Health & Place*, 46(46):328–336, 2017. ISSN 1353-8292. doi: 10.1016/j.healthplace.2016.09.009.

William O. Bearden, Richard G. Netemeyer, and Jesse E. Teel. Measurement of consumer susceptibility to interpersonal influence. *Journal of Consumer Research*, 15:473, 1989. ISSN 0093-5301. doi: 10.1086/209186.

Eran Ben-Elia and Erel Avineri. Response to travel information: A behavioural review. *Transport Reviews*, 35(3):352–377, mar 2015. doi: 10.1080/01441647.2015.1015471.

Bollen. *Structural Equations*. John Wiley & Sons, 1989. ISBN 0471011711. URL `https://www.ebook.de/de/product/3597109/bollen_structural_equations.html`.

G. E. P. Box and D. R. Cox. An analysis of transformations. *Journal of the Royal Statistical Society: Series B (Methodological)*, 26(2):211–243, jul 1964. doi: 10.1111/j.2517-6161.1964.tb00553.x.

Raymond C. Browning, Emily A. Baker, Jessica A. Herron, and Rodger Kram. Effects of obesity and sex on the energetic cost and preferred speed of walking. *Journal of Applied Physiology*, 100(2):390–398, feb 2006. doi: 10.1152/japplphysiol.00767.2005.

Dominik Bucher, Francesca Mangili, Francesca Cellina, Claudio Bonesana, David Jonietz, and Martin Raubal. From location tracking to personalized eco-feedback: A framework for geographic information collection, processing and visualization to promote sustainable mobility behaviors. *Travel Behaviour and Society*, 14:43–56, 2019. ISSN 2214-367X. doi: 10.1016/j.tbs.2018.09.005.

Chiara Calastri, Romain Crastes dit Sourd, and Stephane Hess. We want it all: experiences from a survey seeking to capture social network structures, lifetime events and short-term travel and activity planning. *Transportation*, pages 1–27, 2018. ISSN 0049-4488. doi: 10.1007/s11116-018-9858-7.

Caspar G. Chorus, Eric J. E. Molin, and Bert Van Wee. Use and effects of advanced traveller information services (ATIS): A review of the literature. *Transport Reviews*, 26(2):127–149, mar 2006. doi: 10.1080/01441640500333677.

Caspar G. Chorus, Joan L. Walker, and Moshe Ben-Akiva. A joint model of travel information acquisition and response to received messages. *Transportation Research Part C: Emerging Technologies*, 26:61–77, jan 2013. doi: 10.1016/j.trc.2012.07.002.

Audrey de Nazelle, Mark J. Nieuwenhuijsen, Josep M. Antó, Michael Brauer, David Briggs, Charlotte Braun-Fahrlander, Nick Cavill, Ashley R. Cooper, Hélène Desqueyroux, Scott Fruin, Gerard Hoek, Luc Int Panis, Nicole Janssen, Michael Jerrett, Michael Joffe, Zorana Jovanovic Andersen, Elise van Kempen, Simon Kingham, Nadine Kubesch, Kevin M. Leyden, Julian D. Marshall, Jaume Matamala, Giorgos Mellios, Michelle Mendez, Hala Nassif, David Ogilvie, Rosana Peiró, Katherine Pérez, Ari Rabl, Martina Ragettli, Daniel Rodríguez, David Rojas, Pablo Ruiz, James F. Sallis, Jeroen Terwoert, Jean-François Toussaint, Jouni Tuomisto, Moniek Zuurbier, and Erik Lebret. Improving health through policies that promote active travel: A review of evidence to support integrated health impact assessment. *Environment International*, 37:766–777, 2011. ISSN 0160-4120. doi: 10.1016/j.envint.2011.02.003.

Salvatore Di Dio, Maria La Gennusa, Giorgia Peri, Gianfranco Rizzo, and Ignazio Vinci. Involving people in the building up of smart and sustainable cities: How to influence commuters' behaviors through a mobile app game. *Sustainable Cities and Society*, 42:325–336, 2018. ISSN 2210-6707. doi: 10.1016/j.scs.2018.07.021.

Dick Ettema. Apps, activities and travel: an conceptual exploration based on activity theory. *Transportation*, 45:273–290, 2018. ISSN 0049-4488. doi: 10.1007/s11116-017-9844-5.

European Environment Agency. CO2 emissions from passenger transport. https://www.eea.europa.eu/media/infographics/co2-emissions-from-passenger-transport/view, June 2016.

Brian J Fogg. Persuasive technology: using computers to change what we think and do. *Ubiquity*, 2002(December):5, 2002.

C. Angelo Guevara. Critical assessment of five methods to correct for endogeneity in discrete-choice models. *Transportation Research Part A: Policy and Practice*, 82:240–254, dec 2015. doi: 10.1016/j.tra.2015.10.005.

Stephane Hess and David Palma. Apollo: a flexible, powerful and customisable freeware package for choice model estimation and application. *Journal of Choice Modelling*, Under Review, 2019.

Jerald Jariyasunant, Maya Abou-Zeid, Andre Carrel, Venkatesan Ekambaram, David Gaker, Raja Sengupta, and Joan L. Walker. Quantified traveler: Travel feedback meets the cloud to change behavior. *Journal of Intelligent Transportation Systems*, 19:109–124, 2015. ISSN 1547-2450. doi: 10.1080/15472450.2013.856714.

Uzma Khan and Ravi Dhar. Licensing effect in consumer choice. *Journal of Marketing Research*, 43(2):259–266, may 2006. doi: 10.1509/jmkr.43.2.259.

Jinhee Kim, Soora Rasouli, and Harry Timmermans. Expanding scope of hybrid choice models allowing for mixture of social influences and latent attitudes: Application to intended purchase of electric cars. *Transportation Research Part A*, 69:71–85, 2014. ISSN 0965-8564. doi: 10.1016/j.tra.2014.08.016.

Masashi Kuwano, Makoto Chikaraishi, and Akimasa Fijuwara. Factors that promote personal mobility in relation to the social network in old newtown. *Asian Transport Studies*, 3(1):108–124, 2014.

Michael Maness and Cinzia Cirillo. An indirect latent informational conformity social influence choice model: Formulation and case study. *Transportation Research Part B*, 93:75–101, 2016. ISSN 0191-2615. doi: 10.1016/j.trb.2016.07.008.

mapmywalk.com. Calories calculator. http://www.mapmywalk.com/improve/calorie_calculator/, June 2016.

Italo Meloni, Benedetta Sanjust, Eleonora Sottile, and Elisabetta Cherchi. Propensity for voluntary travel behavior changes: An experimental analysis. *Procedia - Social and Behavioral Sciences*, 87:31–43, 2013. ISSN 1877-0428. doi: 10.1016/j.sbspro.2013.10.592.

Patricia L Mokhtarian and Krishna V Varma. The trade-off between trips and distance traveled in analyzing the emissions impacts of center-based telecommuting. *Transportation Research Part D: Transport and Environment*, 3(6):419–428, nov 1998. doi: 10.1016/s1361-9209(98)00018-2.

James J Murphy, P Geoffrey Allen, Thomas H Stevens, and Darryl Weatherhead. A meta-analysis of hypothetical bias in stated preference valuation. *Environmental and Resource Economics*, 30 (3):313–325, 2005.

Shinichi Nakagawa and Holger Schielzeth. A general and simple method for obtainingR2from generalized linear mixed-effects models. *Methods in Ecology and Evolution*, 4(2):133–142, dec 2012. doi: 10.1111/j.2041-210x.2012.00261.x.

F. B. Norwood and J. L. Lusk. Social desiarability bias in real, hypothetical, and inferred valuation experiments. *American Journal of Agricultural Economics*, (93):528–534, 2011. ISSN 0002-9092. doi: 10.1093/ajae/aaq142.

Robert M. O'Brien. A caution regarding rules of thumb for variance inflation factors. *Quality & Quantity*, 41(5):673–690, mar 2007. doi: 10.1007/s11135-006-9018-6.

Office for National Statistics, National Records of Scotland, and Northern Ireland Statistics and Research Agency. 2011 census aggregate data. techreport, Office for National Statistics and National Records of Scotland and Northern Ireland Statistics and Research Agency, 2016.

Rita Orji, Julita Vassileva, and Regan L. Mandryk. Modeling the efficacy of persuasive strategies for different gamer types in serious games for health. *User Modeling and User-Adapted Interaction*, 24:453–498, 2014. ISSN 0924-1868. doi: 10.1007/s11257-014-9149-8.

Antti Oulasvirta, Tye Rattenbury, Lingyi Ma, and Eeva Raita. Habits make smartphone use more pervasive. *Personal and Ubiquitous Computing*, 16:105–114, 2012. ISSN 1617-4909. doi: 10.1007/s00779-011-0412-2.

Francesco Piras, Eleonora Sottile, Daniele Calli, and Italo Meloni. Automatic data collection for detecting travel behavior: the ipet platform. *Procedia Computer Science*, 134:421–426, 2018. ISSN 1877-0509. doi: 10.1016/j.procs.2018.07.189.

John Pucher, Jennifer Dill, and Susan Handy. Infrastructure, programs, and policies to increase bicycling: An international review. *Preventive Medicine*, 50:S106–S125, 2010. ISSN 0091-7435. doi: 10.1016/j.ypmed.2009.07.028.

R Core Team. *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna, Austria, 2018. URL `https://www.R-project.org/`.

Soora Rasouli and Harry Timmermans. Influence of social networks on latent choice of electric cars: A mixed logit specification using experimental design data. *Networks and Spatial Economics*, 16(16):99–130, 2016. ISSN 1566-113X. doi: 10.1007/s11067-013-9194-6.

Resource Systems Group. rMove. http://rmove.rsginc.com/index.html, 2017.

Lamyae Sardi, Ali Idri, and José Luis Fernández-Alemán. A systematic review of gamification in e-health. *Journal of Biomedical Informatics*, 71:31–48, 2017. ISSN 1532-0464. doi: 10.1016/j.jbi.2017.05.011.

Robert Schlich, Boris Kluge, Sascha Lehmann, and Kay W Axhausen. Durchführung einer 12-wöchigen langzeitbefragung. In *Tagungsband zum 3. Aachener Kolloquium" Mobilität und Stadt", AMUS 2002*, volume 73, pages 141–154. ISB, 2002.

Richard Thaler. Toward a positive theory of consumer choice. *Journal of Economic Behavior and Organization*, 1:39–60, 1980. ISSN 0167-2681. doi: 10.1016/0167-2681(80)90051-7.

Joan L Walker, Emily Ehlers, Ipsita Banerjee, and Elenna R Dugundji. Correcting for endogeneity in behavioral choice models with social influence variables. *Transportation Research Part A: Policy and Practice*, 45(4):362–374, 2011.

Zhenzhen Wang, Sylvia Y. He, and Yee Leung. Applying mobile phone data to travel behaviour research: A literature review. *Travel Behaviour and Society*, 11:141–155, 2018. ISSN 2214-367X. doi: 10.1016/j.tbs.2017.02.005.

Johann Weber, Mojdeh Azad, William Riggs, and Christopher R. Cherry. The convergence of smartphone apps, gamification and competition to increase cycling. *Transportation Research Part F*, 56:333–343, 2018. ISSN 1369-8478. doi: 10.1016/j.trf.2018.04.025.

Paul Weiser, Simon Scheider, Dominik Bucher, Peter Kiefer, and Martin Raubal. Towards sustainable mobility behavior: research challenges for location-aware information and communication technology. *Geoinformatica*, 20:213–239, 2016. ISSN 1384-6175. doi: 10.1007/s10707-015-0242-x.

West Yorkshire Combined Authority. West yorkshire transport strategy evidence base. techreport, West Yorkshire Combined Authority, 2016.

Jeffrey M. Wooldridge. *Econometric Analysis of Cross Section and Panel Data.* MIT Press, Cambridge, MA, second edition, 2010. ISBN 978-0-262-23258-6.