Data Science (2017) 1–22
IOS Press

Supplementary Information

Advancing sustainability: using smartphones to study environmental behaviour in a field-experimental setup

Viktoria Spaiser [a,*] Davide Luzatti [b] and Aliki Gregoriou [c]
[a] *School of Politics and International Studies, University of Leeds, LS6 4RH, Leeds, The United Kingdom*
E-mail: v.spaiser@leeds.ac.uk
[b] *Leeds University Business School, University of Leeds, LS6 4RH, Leeds, The United Kingdom*
E-mail: dadelutz@gmail.com
[c] *School of Computing, University of Leeds, LS6 4RH, Leeds, The United Kingdom*
E-mail: sc14ag@leeds.ac.uk

S1. Bespoke Smartphone Application

Given some technical limitation of using EpiCollect 5 for the pilot study, such as the problem of implementing the field-experimental interventions within the application, work is ongoing on developing a bespoke smartphone application for the study approach described in the main manuscript. However, it may be sensible to consider outsourcing the development of a professional smartphone application to a software company specialised on developing mobile phone applications for a major study, since developing a complex, mutli-functional, high quality mobile phone application is not trivial or at least it is necessary to considerably increase the resources for developing a bespoke application. In this section a brief description of the our work on a bespoke smartphone application will be provided. Further technical details can be obtained from [1].

The bespoke application dataUp, that we have been working on, is a hybrid application, a hybrid between a native application and a web application. Hybrid applications have the advantage that they allow crossplatform development but like native apps, they can be submitted to an app store from where the user can install them and they can take advantage of the many device features available (e.g. taking pictures, etc.) On the other hand, like web applications they rely on HTML being rendered in a browser, though the browser is embedded within the hybrid application. In fact dataUp is in many ways a wrapper for certain parts of the dataUp web application.

The dataUp web application is based on a three-tier architecture with three components or layers: (1) Front-end web tier manages the HTTP requests, (2) the middle tier implements the core functionality and (3) the back-end database stores the data. When a user interacts with the web application or mobile phone application the browser submits a HTTP request to the middle tier and the middle tier then uses the request to retrieve or store data. For the web application framework we chose Python-based Flask web

*Corresponding author.

framework. The advantage of Flask is the variety of features it offers, there are no platform restrictions and it allows for scalability, while providing maximum flexibility. For the design of the web application JQuery, a popular Java Script library was used. Asynchronous JavaScript and XML (AJAX) was used for smooth data retrieval.

In dataUp a non-relational, document-based database was implemented to allow for fast information access and portability. The design is in many ways equivalent to the key value data model, in which values are stored and retrieved via corresponding keys. However in document-based databases the value is semantic and encoded in a standard data exchange formats, such as XML and JSON. Existing documents can be searched for both keys and values. Firebase, a cloud-based NoSQL Database service, that stores data in a JSON tree format, was used to host the application database.

dataUp allows for in-build processing and visualisation of the collected data that is internally coded up numerically in terms of CO2 emissions, i.e. the user responses are automatically associated with respective CO2 emission scores. For the visualisation of the data in the social monitoring intervention group it was decided to use JqPlot, one of the most well known Java Script visualisation libraries, using HTML5 Canvas.

Messaging for the behavioural targeting intervention and for reminders was implemented within the application making use of Firebase Notifications, which is a free service provided by Firebase for user notifications. The context and rules for notifications as well as the targets have to be specified and the Firebase Cloud Messaging service is then responsible to route and deliver those messages to the users through the application. Advanced Python Scheduler within Flask was used for scheduling the messaging, specifically the Cron Style Scheduling that allows date- or time-based scheduling was implemented.

To compile part of the web application as ha ybrid mobile phone application, PhoneGap was used. PhoneGap Build is a service that compiles an application for distinct platforms in the cloud, no native user elements are used, instead a native appearance is simulated, which may be a disadvantage. On the other hand PhoneGap convinces with its simplicity.

Not yet fully implemented in the application is the in-build image taking, scanning of barcode and GPS recording. Moreover the application requires further debugging and testing before it can be fully used in a study. For that reason it was decided to use the widely tested and popular EpiCollect 5 for the pilot study to guarantee smooth and unproblematic data collection, while work on the bespoke smartphone applications continues.

S2. Data Collections

S2.1. Questionnaire Implemented in the App

(1) What is your username? – Open text entry (mandatory).
(2) Please record your start position. – GPS recording.
(3) Please record your destination position. – GPS recording.
(4) Please scan a purchased product (nor grocery shopping). – Barcode scan.
(5) Please scan a purchased product (nor grocery shopping). – Barcode scan.
(6) Please scan a purchased product (nor grocery shopping). – Barcode scan.
(7) What type of transport did you use throughout the day? (multiple choice, mandatory)

- Walking
- Car

- Taxi
- Bicycle
- Bus
- Train
- Plane
- Ferry
- None, stayed at home (jump to question 9, if this answer chosen)

(8) Why did you pick this/these means of transport? (single choice, mandatory)

- Convenience
- Money
- Health
- Environment
- Habit
- Been in company
- No other choice

(9) What did you eat throughout the day? (multiple choice, mandatory)

- Cereals/Muesli
- Pasta/Bread
- Cheese
- Other dairy products
- Meat products (lamb)
- Meat products (beef)
- Meat products (pork)
- Meat products (chicken)
- Fish
- Egg products
- Vegetables/Fruits
- Coffee
- Tea
- Rice
- Grain
- Potatoes
- Soy products

(10) What electronic devices did you use throughout the day? (multiple choice, mandatory)

- Laptop/Notebook
- Mobile Phone
- iPad/Tablet
- External hard drive
- Lamps
- Mircrowave
- Hob/Oven
- Fridge/Freezer

- Washing machine
- Tumble dryer
- Dishwasher
- TV
- Radio/Music player
- Hair dryer
- Shaver
- Air Conditioning/Heater
- Desktop PC

(11) What electronic devices are on (incl. standby mode) throughout the day? (multiple choice, mandatory)

- Desktop PC
- Laptop/Notebook
- External hard drive
- Lamps
- TV
- Radio/Music player
- Air Conditioning/Heater
- None

(12) What waste did you produce throughout the day? (multiple choice, mandatory)

- Paper
- Plastic (not recycled)
- Plastic (recycled)
- General waste
- Biodegradable waste (recycled)
- Electronic waste

(13) What purchases did you make throughout the day? (multiple choice, mandatory)

- Clothes/Shoes
- Clothes/Shoes (eco)
- Clothes/Shoes (second-hand)
- Beverage (incl. alcohol)
- Utensils/Accessoires
- Cosmetic/Hygiene products
- Cosmetic/Hygiene products (eco)
- Small electronic devices
- Small electronic devices (second-hand)
- Book/Newspaper/Magazine (print)
- E-book/E-Newspaper/E-Magazine
- Music CD/Movie DVD
- Music/Movie (MP3, MP4, Streaming, etc.)
- Book/Magazine/CD/DVD (second-hand)
- Nothing

(14) Please take a picture of your electric counter if possible. – Image.

Study participants were encouraged to complete the questionnaire in the evening, starting with the GPS and/or barcode scan recording during the day however.

This questionnaire suffers from various limitations. For instance the user is provided with the option to record one journey only and while this can include several modes of transport, it is possible that users do several smaller journeys throughout the day returning in between home for instance. Furthermore, when users are asked why they have chosen a given means of transport, they can only choose on option, however, in case of them making use of several different transport modes throughout the day, there may be different reasons for each mode of transport. Furthermore, the users are restricted to recording only three barcodes each day. If they buy more then three products, then this is not recorded. Not inquired are also quantities of food or whether the food was organic, regional/seasonal etc., or the duration of usage of electric devices, the type of car, etc.. Thus, the derived CO2 emissions are only approximated on average and are not very precise.

The reason for this and other limitations is the attempt to create a simple app questionnaire that the user can complete as quickly as possible. People are usually not willing to engage with a lengthy questionnaire on their phones on a daily basis and hence are more likely to not complete a questionnaire, skip questions or answer the questions less accurately if it takes them more effort and time to generate the data. It is thus a pay-off between accuracy and user-friendliness. Other limitations, e.g. the barcode scans are also due to the way EpiCollect 5 is implemented. It does not allow to repeatedly collect as many barcode scans as necessary. The same applies to GPS recordings. This is something that should be fixed in a bespoke software.

S2.2. Initial Survey Questionnaire

(1) What is your username? – Open text entry (mandatory).
(2) What is your gender? (mandatory, single choice)

- Male
- Female
- Other

(3) How old are you? – Open text entry (mandatory).
(4) How would you describe your financial situation? (mandatory, single choice)

- Very difficult
- Difficult
- Occasionally difficult
- Overall alright
- Mostly good
- Good
- Very good
- prefer not to answer

(5) What is your attitude to climate change? (mandatory, single choice)

- It is one of the most serious problems that humanity is facing today.
- It is a quite serious problem.
- It is a serious problem, but I am not sure anything can be done about it.

- I think the problem of climate change is overstated.
- I don't really know much about climate change.
- I don't think climate change is real.

(6) What is your attitude to recycling? (single choice)

- I always recycle.
- I try to recycle when possible.
- I am not sure about what can be recycled.
- I don't think recycling is important.
- I don't have the possibility to recycle where I live.

(7) What is your attitude to buying ecological products? (single choice)

- I buy ecological products whenever possible.
- I often buy ecological products.
- Ecological products are too expensive.
- I don't really care.

The survey was implemented online on http://en.q-set.co.uk.

S2.3. Final Survey Questionnaire

(1) What is your username? – Open text entry (mandatory).
(2) Please describe your experience as a study participant of this study. – Open text entry.
(3) Did the participation in this study raise your awareness of your ecological footprint? (single choice, mandatory)

- Yes, absolutely.
- Yes, to some extent.
- Only little.
- Rather not.
- Not at all.
- Not sure.

(4) What did you like about taking part in this study? – Open text entry.
(5) What were the things you did not like about participating in this study? – Open text entry.
(6) How can the data-collection application be improved? – Open text entry.
(7) Was the compensation for participating in this study fair? (single choice)

- Yes
- A bigger compensation should have been given
- A smaller compensation would have been sufficient

(8) Would you participate in such a study if the study would stretch over four rather than two weeks? (Compensation would be increased accordingly) (single choice, mandatory)

- Yes
- No
- Don't know

The survey was implemented online on http://en.q-set.co.uk.

S2.4. Notifications

The reminder notification was: “Hello (username), this is a kind reminder please to generate and upload your daily data via the EpiCollect 5 App. It should take only a few minutes. If you have any questions or encounter any problems, please contact Dr Spaiser: v.spaiser@leeds.ac.uk. Many thanks!”

Notifications in the behavioural targeting group were individualised, based on the users environmental performance. For instance, if the CO2 emissions were particularly high, due to consumption of certain meat products, then the user would receive the following message: “Hello (username), your environmental impact score on the food dimension was somewhat higher yesterday. Meat products in particular from beef and lamb are very problematic from an environmental point of view. You can reduce your ecological footprint considerably by avoiding eating beef and lamb.”

Or if the user had higher than average CO2 emissions in the energy usage dimension, they would receive a notification saying: “Hello (username), how about reducing your energy bill while at the same time protecting the environment? You can start by thinking which devices are on, even though you don’t actually actively use or need them, like the (devices, that the user left on throughout the day), being on or on standby throughout the day even though you do not use them.”, if the higher CO2 emissions were due to devices being on all day, otherwise the following message: “Hello (username), it is difficult to save energy with all the devices we are using every day, but usually, there is room for some energy saving. Have you for instance considered drying your clothes on a clothes horse/washing line outside if it is not raining, instead of using the tumble dryer?".

If the CO2 emissions were high in the transport dimension due to car usage, then the study participant would receive one of these two notifications to avoid repetition if the message needed to be sent out at least twice on two consecutive days: “Hello (username), cars can be a convenient transport means, but did you know that you are affected more by air pollution produced by cars when driving a car in comparison to when you are cycling or walking? You can reduce your ecological footprint and reduce air pollution by cycling, walking or using the bus or train.” or “Hello (username), your environmental impact score on the transport dimension is slightly higher than average. It’s the car you are using. You can reduce your CO2 emissions considerably if you choose to cycle, walk or take the bus or train instead.” Or if a user used a taxi, which resulted in higher CO2 emissions, they would receive a similar messages: “Hello (username), your environmental impact score on the transport dimension was somewhat higher yesterday. It’s the taxi you were using. You can reduce your CO2 emissions considerably if you choose to cycle, walk or take the bus or train instead." or “Hello (username), taxis can be a convenient transport means, but did you know that you are affected more by air pollution produced by cars when travelling in a car in comparison to when you are cycling or walking? You can reduce your ecological footprint and reduce air pollution by cycling, walking or using the bus or train." In the case of flying, study participants would receive a notification saying: “Hello (username), we all know that flying is not environmentally friendly and that it contributes enormously to our ecological footprint, but sometimes it is just unavoidable. However, at other times other options are available and if not there is the voluntary option to offset your carbon footprint."

When high CO2 emissions levels were reached in the waste dimension du to non-recycled plastic, the study participant would receive an email saying: “Hello (username), did you consider buying products that use less packaging or buying unpackaged products like vegetables and fruits, not putting them in a plastic bag to reduce unnecessary waste of plastic? You could also try to buy products that are packaged in recyclable packaging to reduce the ecological impact of waste.”

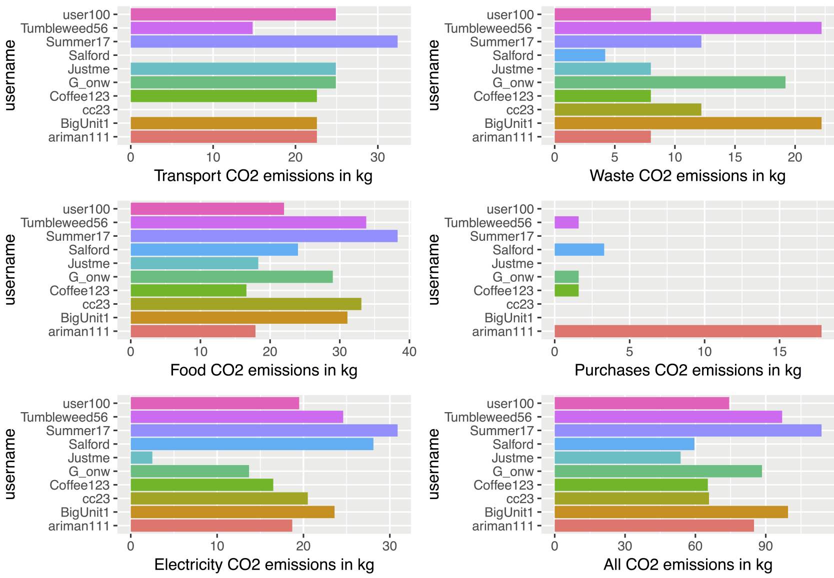

Fig. 1. Exemplary image that study participants in the social monitoring treatment group would receive

If a study participant did rather well on the previous day, having below average CO2 emissions on all dimensions, they would receive the following message: “Hello (username), your environmental performance was good yesterday, keep it up!” And if a user failed to upload data the previous day, they would receive the following notification: “Hello (username), sorry, today we cannot send you any individual advice because we did not receive any data from you yesterday.”

In the social monitoring treatment group, the message that study participants would receive, would say “see your environmental performance in comparison to others in your group on (date)" in the email subject field and the email body would contain an image like the one in Figure 1.

All notifications were sent out from the pilot study email account ecosmartpilot@gmail.com, always at the same time, treatment notifications at 5pm, reminder notifications at 10pm.

S3. Supplementary Results

S3.1. Descriptives

As mentioned in the main text, from the N=20 study participants, 12 were female and 8 male. Moreover 13 were students and 7 had a professional background. The age distribution can be seen in Figure 2, it ranged between 18 and 43 years, with the mean of 25.7 and standard deviation of 7.23. Most study participants assessed their financial situation as good, no one responded being in “very difficult" or “difficult" financial situation, though three respondents said that their financial situation is “occasionally difficult". The median response was “mostly good". When it comes to climate change attitudes, the vast majority of 16 said that they thought that climate change is the most serious problem humanity is facing today. Respectively 2 said that it is a “quite serious" and “serious, but that they were not sure what to do about it". No one thought that the problem of climate change is overstated, or that climate change is not real and no one said they did not know much about climate change. With respect to recycling, all study participants claimed to recycle either always (8) or whenever possible (12). No study participants claimed not to be sure what can be recycled or not having the possibility to recycle where they live. Equally no one thought recycling is not important. Study participants were also asked in the initial survey what their attitude to buying ecological products is. Figure 2 shows that many respondents (8) thought that “ecological products are too expensive", while 7 claimed to “often buy ecological products" and 5 even to “buy ecological products whenever possible". No study participant said they “don’t really care". These attitudes results show that the study participants displayed strong pro-environmental preferences and attitudes. But are these attitudes also visible in their everyday behaviour? As the later analyses presented here and in the main paper show, this is not necessarily the case. In the post-study survey finally the study participants were also asked whether they thought that the participation in the study raised their ecological awareness. Figure 2 shows that the study participants varied in their responses to this question, with the majority (17) saying that the study participants raised their ecological awareness “only a little" (10), “to some extent" (6) or “absolutely" (1). Three study participants said that the study participation did “rather not" (2) or “not at all" (1) raised their ecological awareness. The median was “only a little". Given however that it appears that most study participants had already a high ecological awareness prior to the study as the attitude responses in Figure 2 suggest, it is interesting to note, that most of them still felt that the study had some effect on their “ecological awareness". However, again if it comes to actual behavioural change, then the study effect is much less clear, as will be shown in the following and as is shown in the main paper.

The actual environmental behaviour can be derived from the data collected through the smartphone application app over 14 days. Figure 3 shows the CO2 emissions trends over the two weeks of data collections for some exemplary study participants. We see considerable fluctuations. Only few display regular patterns, e.g. penguin89 seems to have mostly constant transport CO2 emissions, which indicates a commuting behavior with a particular transport mode. Indeed penguin89 was one of the professionals in the datasets. Students on the other hand seem to vary much more, which is partly due to the holiday season.

Looking at the means (see Figure 4) we can see again fluctuating patters that however reveal some important dynamics. For instance in the transport dimension we see that the fluctuations are much stronger in the second week, with overall higher CO2 emissions in the second week. As noted in the main paper, many more study participants did fly in the second week comparing to the first week. No clear average pattern is discernible for the food, waste and purchases dimension, though there seems to be an average increasing tendency in the food dimension. The electricity dimension seems to suggest a general average decrease, though interrupted by some increases in between.

Most study participants did not take images of their electric meter, thus data on the actual electricity usage is scarce. And as reported in the main paper there are serious issues with retrieving the data from the images of the electric meter counter. For those six study participants who took images of their electric meter counter, we however manually retrieved the data from the images and added it to the existing data. Figure 5 shows the electricity usage in kWh (kilo Watt per hour) for the six study participants who provided their data. We would naturally expect that the curve is always increasing, so the measure of interest is the rate of increase and here we do see some variation, though not necessarily suggesting a clear pattern. Some curves (e.g. cc23, penguin89, Tumbleweed56) seem to flatten, but then depict a higher rate of increase again. Since the data on actual electricity usage was scarce, it was not included in the calculations of the electricity CO2 emissions for each individual. This is however something that should be ideally pursued in an actual study.

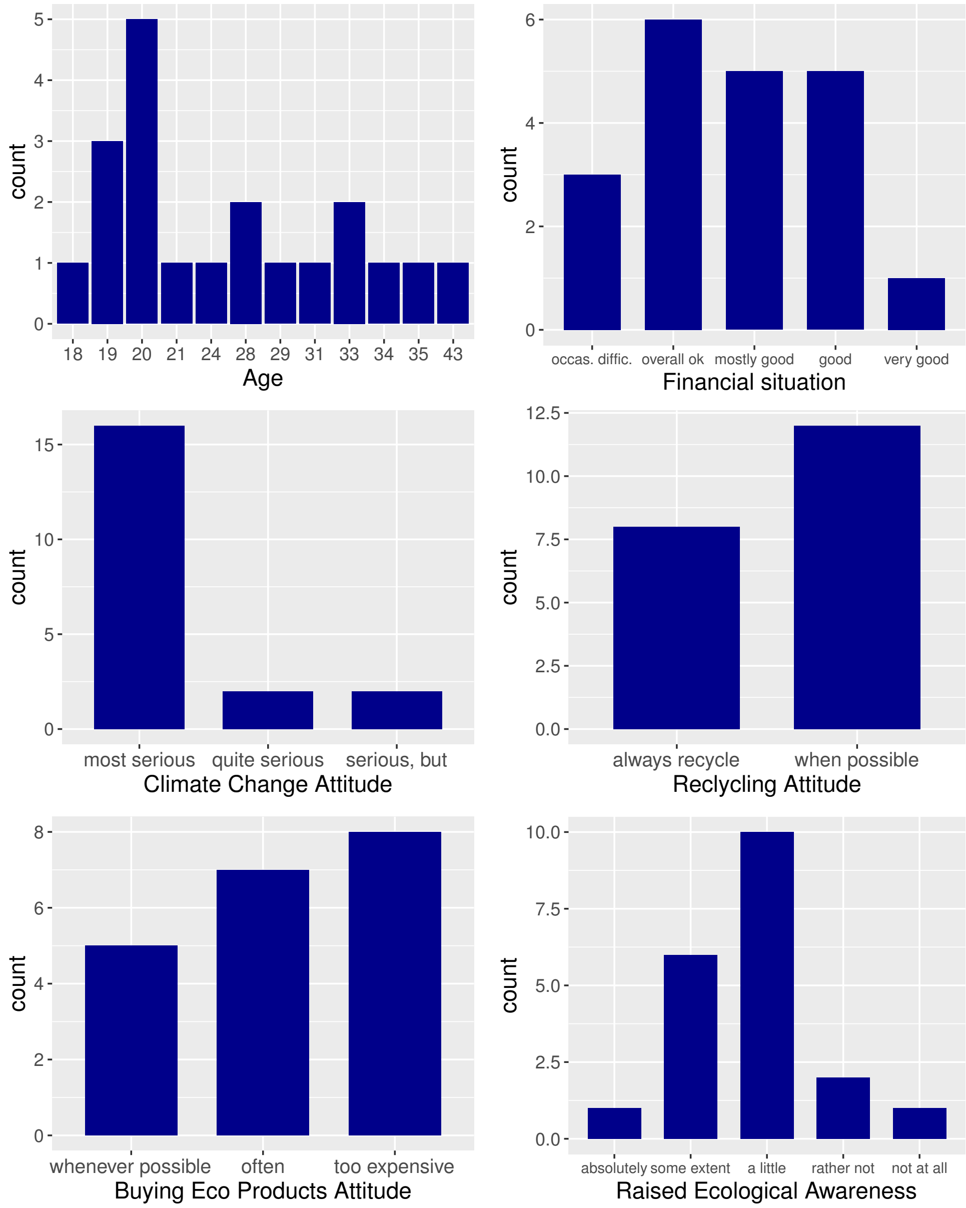

Fig. 2. Bar charts for various variables from the initial and final survey. Age ranged between 18 and 43, with the mean of 25.7 and standard deviation of 7.23. Financial situation ranged between 1 “very difficult" and 7 “very good". The median of this variable was 5, “mostly good". Climate change attitudes ranged between 1 “most serious problem" and 6 “climate change is not real". The median of this variable was 1, “most serious". Recycling attitude ranged between 1 “I always recycle" and 5 “I don’t have the possibility to recycle where I live". The median of this variable was 2, “I try to recycle when possible". Buying ecological products attitude ranged from 1 “I buy ecological products whenever possible" to 4 “I don’t really care". The median of this variable was 2 “I often buy ecological products". The Raised ecological awareness final survey variable ranged between 1 “Yes, absolutely" and 5 “Not at all". The median of this variable is 3 ‘only little".

Fig. 3. CO2 emissions trends for 8 exemplary study participants. The field-experimental intervention phase comprised of the days 8 to 14

Data obtained from barcode scans was again not included in the calculations of CO2 emissions, because of scarce and incomplete data provision by the study participants. Moreover some of the barcodes could not be deciphered with existing, accessible barcode datasets. Where data was available and could be deciphered however, the answers given to the purchases question were compared with the information obtained from the barcode scans. The answers provided by the participants were usually accurate and could be confirmed in most cases through barcode scans, though barcode scan data gives, if available and decodable, much more detailed information, that could be further exploited (e.g. precise item, quantity, price, etc.). Given this information richness, barcode scans should be preferred over survey responses in an actual study, though that would require participants' collaboration and commitment.

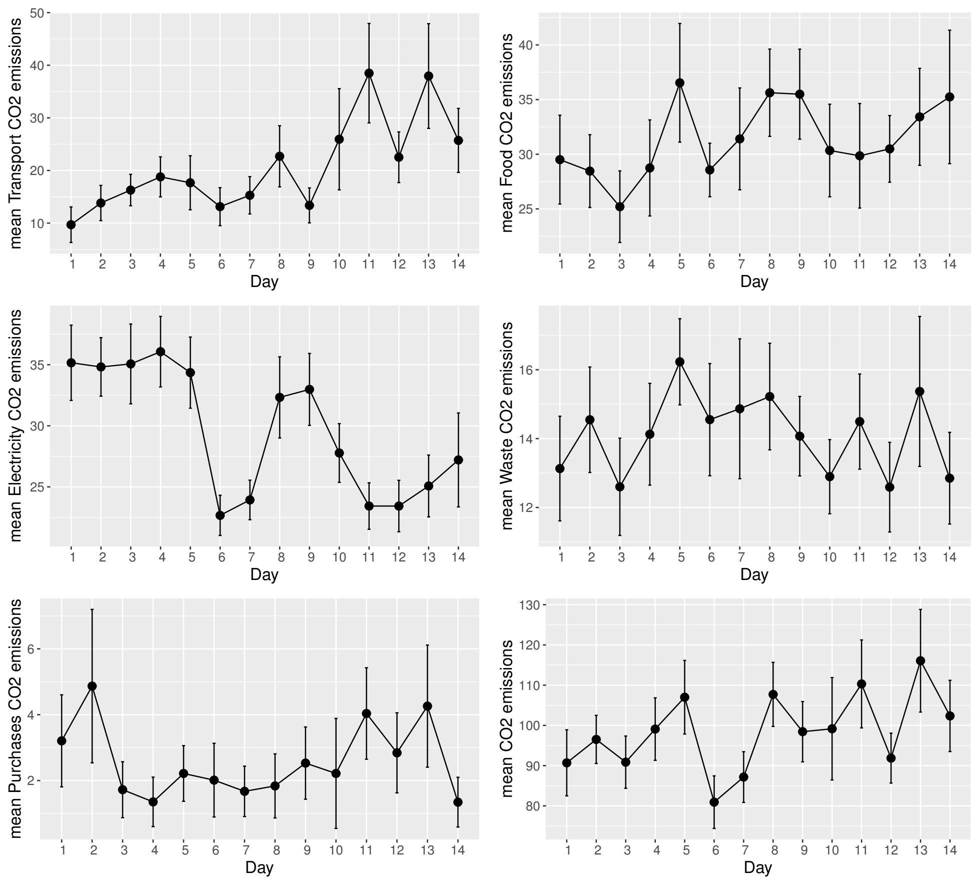

Fig. 4. Mean CO2 emissions, incl. standard error bars over the 14 days of data collection

S3.2. Treatment Effects

For the transport dimension, we can see the CO2 emissions descriptives before and after treatment for each study participant in Figure 6. The picture is rather inconsistent, in some cases the transport CO2 emissions were higher in the second week (after intervention), in other cases it was lower and again in others almost the same comparing to the first week (before intervention). There is moreover considerable variation within individuals and in some cases notable outliers.

The paired t-test shows that there is a significant difference between the mean transport CO2 emission before and after the field-experimental intervention ($t = -3.14$, $p = 0.01$, CI: [-21.20, -4.26]) and the mean difference (-12.73) points clearly in the opposite direction of the expected treatment effect. The aggregate mean CO2 emissions in the second week (after intervention) (28.34, sd = 17.84) is considerably higher than in the first week (15.61, sd = 9.38), though the variance is higher in the second week too. As already mentioned in the main text, this is due to the fact that many study participants were flying in the second week (7 out of 20 comparing to 1 out of 20 in the first week). Given that people tend to make their travel plans and to book their flights some considerable time in advance, it is safe to assume that the sudden rise of flights in the second week (and hence the rise in mean transport CO2 emissions) was not due to but rather despite the field-experimental treatment.

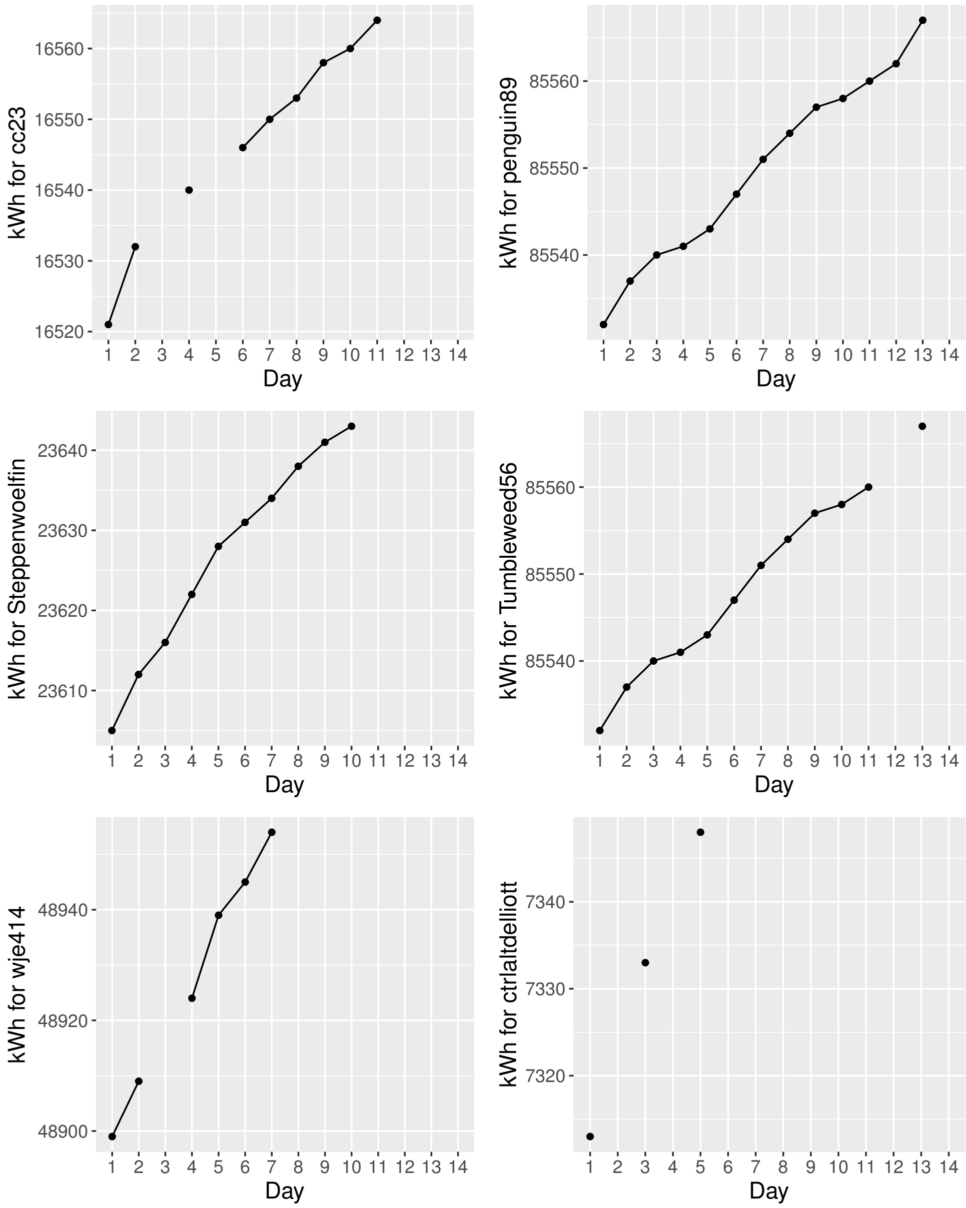

Fig. 5. Electricity usages based on electric meter data for six study participants over 14 days of data collection

For the food dimension, we can see the CO2 emissions descriptives before and after treatment for each study participant in Figure 7. Here too the picture is rather inconsistent, in quite a few cases the food CO2 emissions were higher in the second week (after intervention), in other cases it was lower and again in others almost the same comparing to the first week (before intervention). There is moreover considerable variation within individuals and in some cases notable upper outliers.

The paired t-test shows that there is a slightly significant treatment effect ($t = -2.32$, $p = 0.03$, CI:

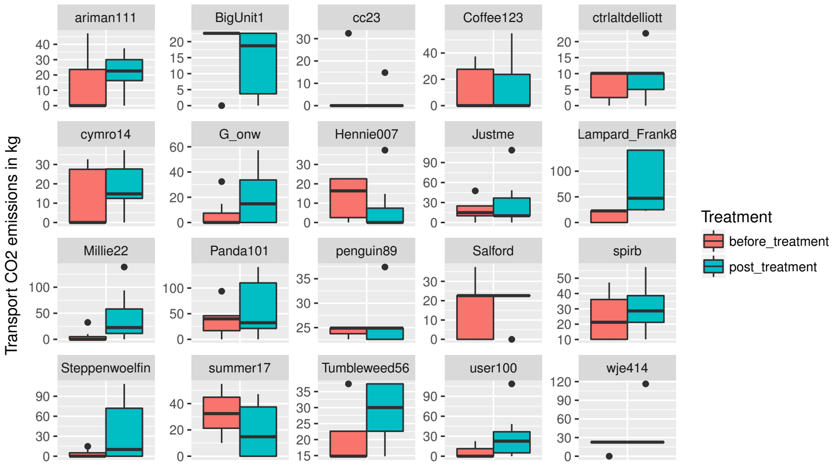

Fig. 6. Boxplots for the 20 study participants showing their transport CO2 emissions before and after the field-experimental treatment.

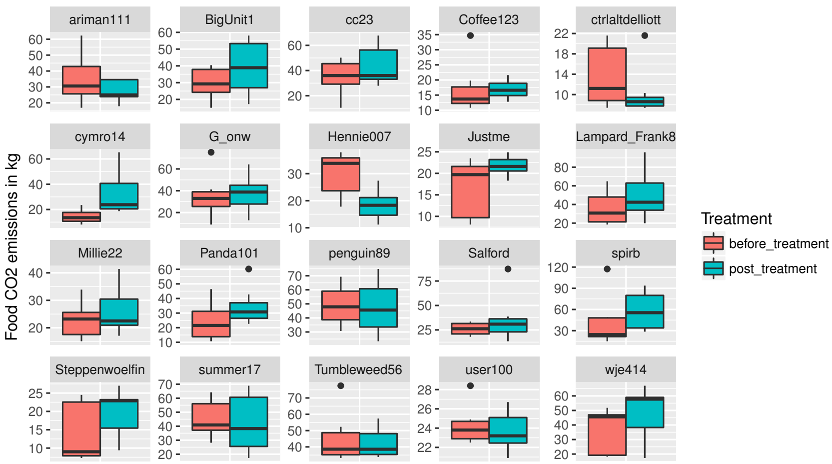

Fig. 7. Boxplots for the 20 study participants showing their food CO2 emissions before and after the field-experimental treatment.

[-7.91, -0.41]), which, with the mean difference of -4.16 even points in the opposite direction, that is the aggregate mean food CO2 emissions in the second week (after intervention) (33.92, sd = 13.03) is somewhat higher than in the first week (29.76, sd = 11.39).

Figure 8 shows the electricity CO2 emission descriptives before and after treatment for each study participant. The figure shows that at least in this dimension we see in many cases, that the electricity usage was lower after the field-experimental treatment comparing to levels of the first week (before treatment). In some cases there is considerable variation and again notable outliers.

This impression is confirmed by the paired t-test, which shows that there is a significant positive treatment effect (t = -2.87, p = 0.01, CI: [1.16, 7.18]) with the mean difference of 4.15. In this case

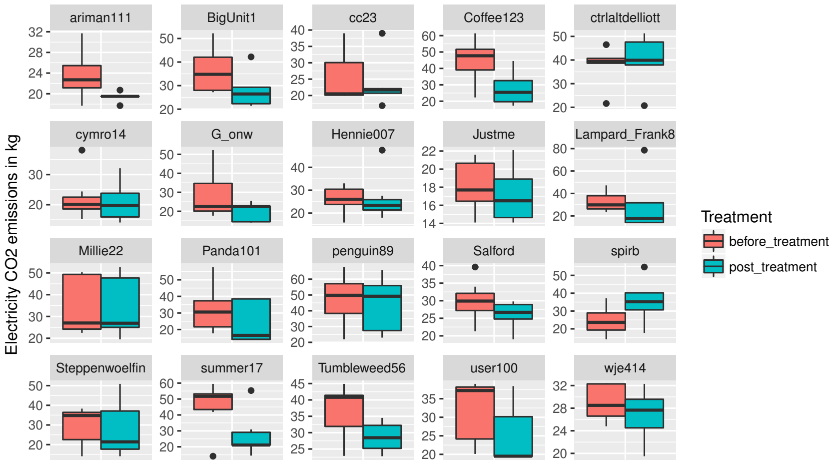

Fig. 8. Boxplots for the 20 study participants showing their electricity CO2 emissions before and after the field-experimental treatment.

thus, the aggregate mean CO2 emissions in the second week (after intervention) (27.73, sd = 6.80) is significantly lower than in the first week (31.88, sd = 7.86). Whether this is indeed due to the treatment is rather difficult to judge based given the limitations of such a small pilot study. It could be as well the side-effect of increased traveling. When traveling people tend to use fewer electronic devices throughout the day and as mentioned earlier, we know that many more study participants were traveling in the second comparing to the first week.

In the case of Waste CO2 and Purchases CO2 the picture becomes inconsistent again as Figures 9 and 10 show. Study participants varied quite significantly in their waste CO2, which may be partly due to inaccurate reporting. And in the case of purchases, there is again considerable variation between and within individuals, with some individuals (e.g. Henni007) having made no purchases at all during the two weeks, which could be also due to underreporting.

The results from the paired t-tests are similarly inconclusive. For the waste dimension the test shows that there is no significant treatment effect (t = -0.16, p = 0.88, CI: [-2.00, 1.72]), and although the mean difference (-0.14) points slightly in the opposite direction, the confidence interval contains a negative and a positive value, which shows the inconclusiveness. The aggregate mean CO2 emissions in the second week (after intervention) (14.38, sd = 5.19) is almost the same as in the first week (14.24, sd = 4.25). Similarly, there is no significant treatment effect (t = -0.36, p = 0.73, CI: [-2.00, 1.41]) for the purchases dimension. The mean difference (-0.29) points again slightly in the opposite direction, but the confidence interval contains a negative and a positive value. The aggregate mean CO2 emissions in the second week (after intervention) (2.99, sd = 3.03) differs only slightly from the value in the first week (2.70, sd = 2.76).

S3.3. Repeated Measures ANOVA

The Repeated Measures ANOVA results for the different environmental behaviour dimensions are displayed in Table 1 and 2. The model fit of the random effect model for the transport dimension CO2 emissions is not very good, in fact the results suggest that accounting for random effects does not improve the model in comparison to a fixed effect model (see Table 1). Username could be therefore removed as

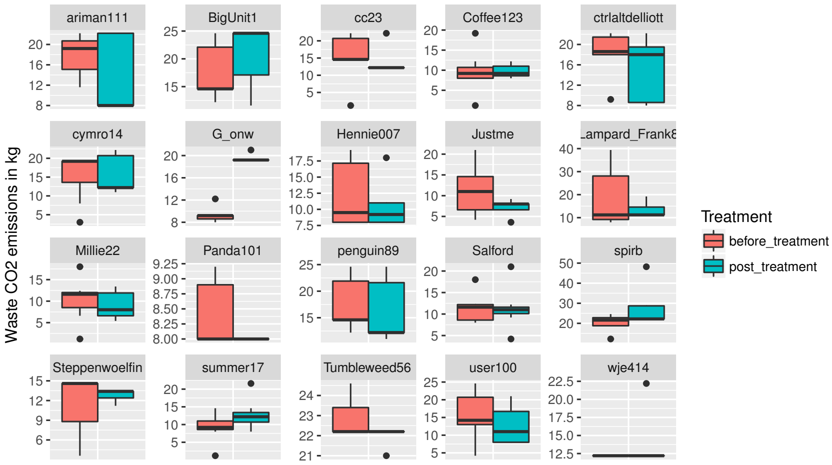

Fig. 9. Boxplots for the 20 study participants showing their waste CO2 emissions before and after the field-experimental treatment.

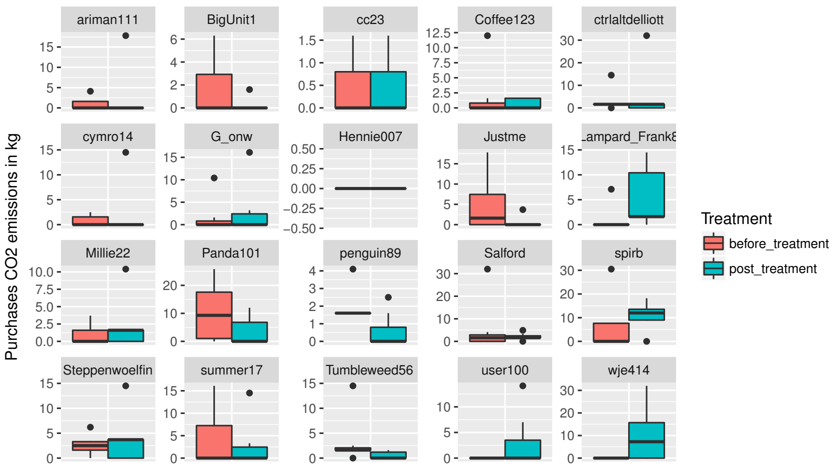

Fig. 10. Boxplots for the 20 study participants showing their purchases CO2 emissions before and after the field-experimental treatment.

a random variable. A model without random effects, but with autocorrelation included seems therefore of a better fit. A comparison with null models confirms that that random effect model is not very strong. However, even when comparing the null models with a model that does not include random effects (only autocorrelation), suggests that even the model without random effects is not of great fit to the data (vs null model LL-diff: -4.04, Chi-Square: 8.08, p = 0.09; vs null model with random effect LL-diff: 1.97, Chi-Square: 3.93, p = 0.27). The treatments thus seem to have essentially no predictive power when it comes to transport CO2 emissions. Nevertheless, if any treatment has any effect, then it seems to be rather the social monitoring treatment rather than the behavioural targeting (see Figure 11).

Table 1

Random effect (r.e.) models vs. fixed effect (f.e.) model, based on Log Likelihood (LL), AIC, BIC and the Log Likelihood Ratio test

Model	Transport			Food			Electricity			Waste			Purchases		
	LL	AIC	BIC	LL	AIC	BIC	LL	AIC	BIC	LL	AIC	BIC	LL	AIC	BIC
r.e. model	-634.01	1282.02	1301.99	-550.01	1114.02	1133.99	-498.60	1011.21	1031.17	-401.53	817.07	837.03	-415.68	845.35	865.32
f.e. model	-636.02	1282.04	1296.30	-568.22	1146.44	1160.70	-503.96	1017.92	1032.18	-428.35	866.70	880.96	-415.96	841.89	856.15
L-Ratio test	4.02, p = 0.13			36.42, p < 0.01			10.71, p = 0.005			53.63, p < 0.01			0.54, p = 0.76		

Table 2

Random effect model vs. null model and null model with random effects (r.e.), based on McFadden Pseudo R Square (McF), Cox/Snell Pseudoe R Square (C/S) and Nagelkerke Pseudo R Square (N) and Log Likelihood difference test

Model	Transport			Food			Electricity			Waste			Purchases		
	McF	C/S	N	McF	C/S	N	McF	C/S	N	McF	C/S	N	McF	C/S	N
vs. null model	0.01	0.08	0.08	0.03	0.22	0.22	0.03	0.19	0.19	0.06	0.34	0.34	0.01	0.05	0.05
vs. r.e. null model	0.01	0.06	0.06	0.01	0.08	0.08	0.02	0.14	0.14	0.02	0.11	0.11	0.01	0.04	0.04

Log Likelihood (LL) difference test

Model	Transport	Food	Electricity	Waste	Purchases
vs. null model	LL-diff: -5.82, Chi-Sq: 11.64, p = 0.04	LL-diff: -16.77, Chi-Sq: 33.53, p < 0.01	LL-diff: -13.98, Chi-Sq: 27.95, p < 0.01	LL-diff: -27.00, Chi-Sq: 53.99, p < 0.01	LL-diff: -3.16, Chi-Sq: 6.32, p = 0.28
vs. r.e. null model	LL-diff: -3.75, Chi-Sq: 7.50, p = 0.11	LL-diff: -5.58, Chi-Sq: 11.16, p = 0.02	LL-diff: -9.74, Chi-Sq: 19.49, p < 0.01	LL-diff: -7.41, Chi-Sq: 14.81, p = 0.01	LL-diff: -2.82, Chi-Sq: 5.64, p = 0.23

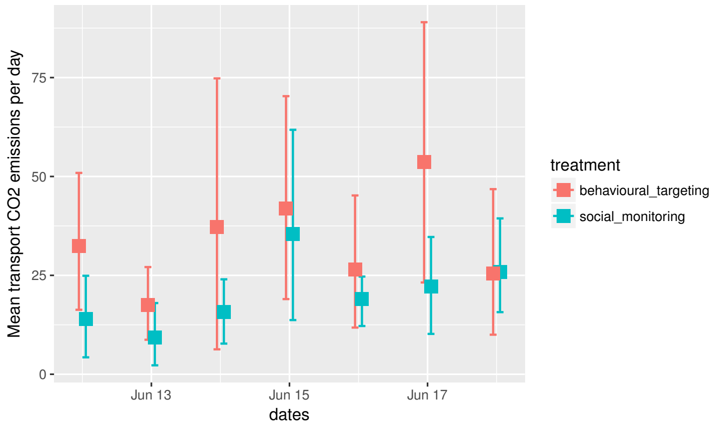

Fig. 11. Interaction plot for transport dimension shows the natural mean of each treatment*date combination along with the confidence interval of each mean with percentile method. Least-Square means for CO2 emissions in the behavioural targeting group were estimated to be 34.28 (se: 5.22, CI: [21.60, 46.96]), for social monitoring 20.19 (se: 5.12, CI: [7.70, 32.67]).

The model fit of the random effect model for the food dimension CO2 emissions is again a reasonable model. As the results in Table 1 show the random effect model is of better fit to the data than a fixed effect model. A comparison with both null models (see Table 2) confirms that the random effect model is indeed of good fit. Accounting for the treatments thus indeed increases the predictive power of the model for food CO2 emissions. On the other hand, it is less clear in the case of the food dimension, which of the two treatment has the strongest effect on behavioural change in terms of reducing CO2 emissions (see Figure 11).

Already the paired t-test suggested that if the treatments had any effect on environmental behaviour over the short period of data collection, then it was in the electricity dimension. This is further supported by the results of the repeated measures ANOVA, that shows a significant treatment effect (Analysis

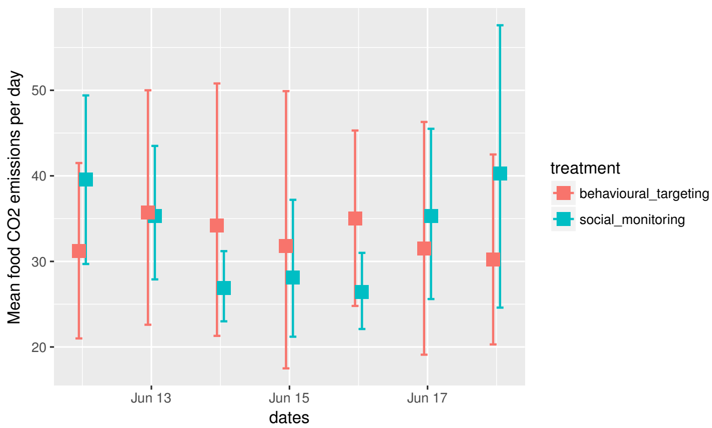

Fig. 12. Interaction plot for food dimension shows the natural mean of each treatment*date combination along with the confidence interval of each mean with percentile method. Least-Square means for CO2 emissions in the behavioural targeting group were estimated to be 35.08 (se: 4.56, CI: [24.01, 46.15]), for social monitoring 34.44 (se: 4.54, CI: [23.38, 45.50]).

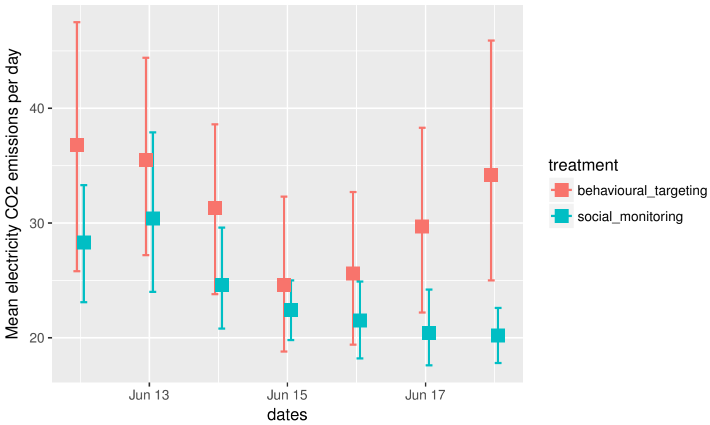

Fig. 13. Interaction plot for electricity dimension shows the natural mean of each treatment*date combination along with the confidence interval of each mean with percentile method. Least-Square means for CO2 emissions in the behavioural targeting group were estimated to be 31.78 (se: 1.94, CI: [27.06, 36.50]), for social monitoring 24.18 (se: 1.92, CI: [19.50, 28.86]).

of Deviance Chi-Square 7.75, $p = 0.005$ for treatment), which was not the case for the two previous dimensions transport and food.The model fit of the random effect model for the electricity dimension CO2 emissions is satisfactory. As the results in Table 1 show the random effect model is of better fit to the data than a fixed effect model. A comparison with both null models (see Table 2) confirms that the random effect model is indeed of good fit. Accounting for the treatments thus indeed increases the predictive power of the model for electricity CO2 emissions considerably. And as Figure 13 and the Least Square Means analyses show, the social monitoring treatment has a stronger and more consistent effect on behavioural change in terms of electricity-base CO2 emissions reductions comparing to behavioural targeting treatment.

The model fit of the random effect model for the waste dimension CO2 emissions is a sensible model. As the results in Table 1 show the random effect model is of better fit to the data than a fixed effect model. A comparison with both null models (see Table 2) confirms that the random effect model is indeed of good fit. Accounting for the treatments thus indeed increases the predictive power of the model for waste CO2 emissions. On the other hand, it is less clear in the case of the waste dimension as was already the case in the food dimension, which of the two treatments has the strongest effect on behavioural change in terms of reducing CO2 emissions (see Figure 14). There is hardly any difference between the two treatments.

The model fit of the random effect model for the purchases dimension CO2 emissions is similarly to the model for the transport dimension not very good, in fact, the results suggest that accounting for random effects does not improve the model in comparison to a fixed effect model (see Table 1). Username could be therefore removed as a random variable. But, even a model without random effects, but with autocorrelation included seems not to fit the data much better (vs. random effect model L-ratio: 0.52, p=0.47). A comparison with null models confirms that that the model is not very strong. The treatments thus seem to have essentially no predictive power when it comes to purchases CO2 emissions, which seem not follow a predictive pattern in general potentially due to insufficient data. Nevertheless, if any treatment has any effect, then it seems to be rather the social monitoring treatment rather than the behavioural targeting (see Figure 15).

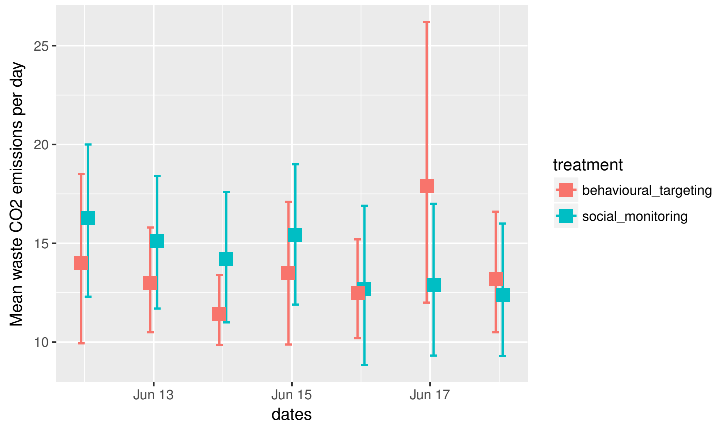

Fig. 14. Interaction plot for waste dimension shows the natural mean of each treatment*date combination along with the confidence interval of each mean with percentile method. Least-Square means for CO2 emissions in the behavioural targeting group were estimated to be 14.61 (se: 1.77, CI: [10.33, 18.89]), for social monitoring 14.54 (se: 1.76, CI: [10.25, 18.83]).

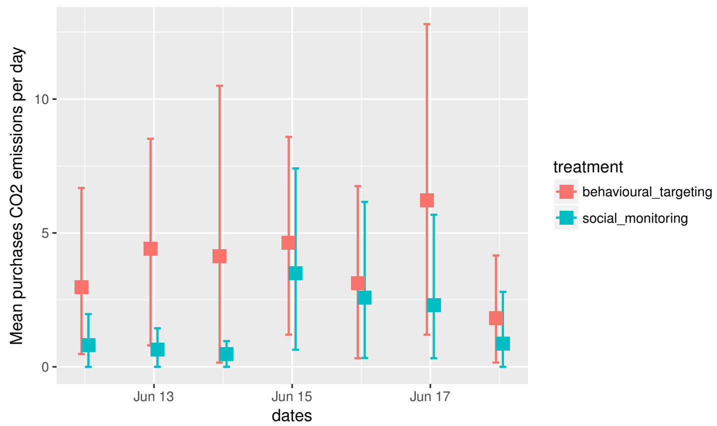

Fig. 15. Interaction plot for purchases dimension shows the natural mean of each treatment*date combination along with the confidence interval of each mean with percentile method. Least-Square means for CO2 emissions in the behavioural targeting group were estimated to be 4.06 (se: 0.79, CI: [2.13, 5.99]), for social monitoring 1.63 (se: 0.77, CI: [0, 3.52]).

S3.4. Gaussian Processes Choice-models

In the main manuscript the utility function of transport modes was discussed mostly with respect to CO2 emissions as a choice characteristic and very briefly with respect to the level of independence that a transport mode allows for. Here we will mostly look into the transport costs as a choice characteristic. However, we can briefly confirm that the utility function pattern for transport modes based on CO2 emissions and with respect to distance look very similar for older participants, either well off or poor (see Figure 16, upper two panels) as they did for younger participants in the main manuscript.

When it comes to the travel costs, the results suggest that study participant do indeed make a transport choice taking costs into account. As Figure 16 shows, older and younger poorer participants have a

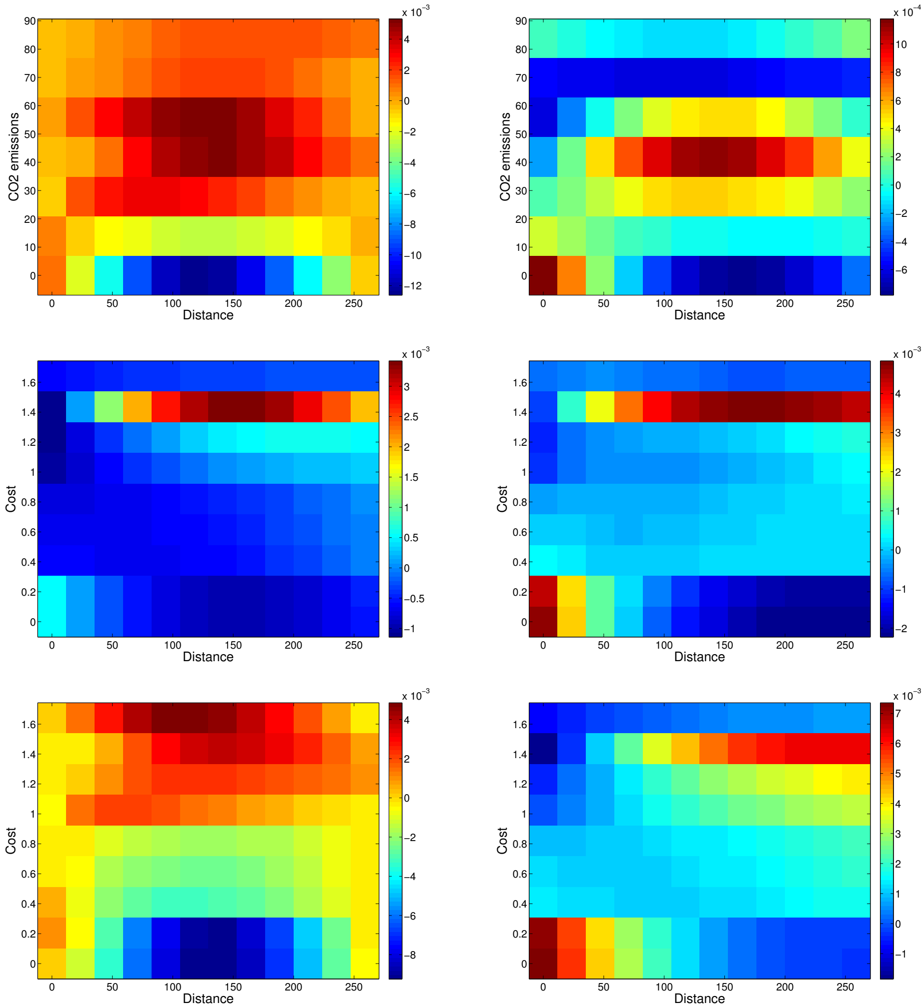

Fig. 16. Heat plots displaying the utility function for transport modes based on CO2 (upper two panels) and costs as choice characteristics. The colour bar shows the utility scale, with redder colours indicating a positive utility and bluer colours none or even negative utility. The upper two panels show transport mode preferences with respect to travel distance for older and financially well (left) or poor (right) participants (a similar pattern emerges for younger participants, see main manuscript). The two panels in the second row show the effect of distance for older but poorer participants (left) and younger poorer participants (right). The two panels in the bottom row show the effect of distance for older but well-off participants (left) and younger well-off participants (right).

similar preference for a certain transport mode of a certain cost once the distance does not allow to take lower cost transport modes, for which the younger have a slightly stronger preference. However this might be an artefact of lack of data, since the financially poorer study participants were rather young. We see that the range of choices of transport modes of various average costs is higher for more well-off older participants, but generally longer distance require higher-costs transport modes. The range of transport modes in terms of costs seems to be more narrow for younger, well-off study-participants, in fact their utility patterns are very similar to younger poorer study participants. For shorter distances they clearly prefer low to no-cost transport modes, for longer distances higher costs are accepted, but they never display a preference for the most expensive transport modes as the older participants do, no matter what distance.

References

[1] A. Gregoriou *Application for Environmental Behaviour Data Collection and Intervention*, Final Year Undergraduate Project, School of Computing, Faculty of Engineering, University of Leeds, 2017, https://vlebb.leeds.ac.uk/bbcswebdav/orgs/SCH_Computing/FYProj/reports/1617/GREGORIOU17-FPR.pdf.