Data Science (2018) 1–29
IOS Press

Supplementary Information

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

Viktoria Spaiser [a,*] Davide Luzzatti [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: alikigreg@gmail.com

S1. Smartphone Application

For the pilot study the free EpiCollect 5 Mobile and Web Application (https://five.epicollect.net), developed by Imperial College London, was used for data collection purposes [1, 2]. The platform allows for creation of project-specific smartphone applications with respective inquiry forms on the EpiCollect 5 website and then publish these through the EpiCollect 5 mobile phone application, that operates on iOS and Android smartphones. After installing the EpiCollect 5 application on their phones, users can then select and load the published project through the app. Following that they are ready to generate data online and offline through their mobile devices any time and as often as necessary, save it and upload it to the EpiCollect 5 server the next time they are online. EpiCollect 5 gives users full control over their data; they have to explicitly upload the data. The researcher is able to view and download the user-uploaded data from the server either in JSON or CSV format. EpiCollect 5 allows to collect the following data types: (1) simple or multiple choice questions or text entries, (2) GPS coordinates, (3) images, (4) videos, (5) audio and (6) barcodes. In the pilot study (1), (2), (3) and (6) were used for data collection.

EpiCollect 5 is not designed for experimental research, i.e. there is no way to issue treatments directly through the app, like sending users messages (this was possible in the previous EpiCollect version though) or visualizing their (and others') data entries (e.g. environmental performance) through the app. In the pilot study, notifications were sent to users' email addresses in both treatment cases, with messages containing either individualized advice on how to reduce ecological footprint based on individual past data or with messages containing visualizations of their own and others' environmental performance for social monitoring purposes. A bespoke smartphone application is currently work in progress and would

*Corresponding author.

allow for sending messages directly and automatically (incl. reminders to upload data) to users through the app and to visualize environmental performance within the app (see further details in section S1.1).

The pilot study showed that collecting electric meter data through image taking is problematic, because study participants take pictures of various quality and perspective. Each electric meter is moreover different in shape and functionality, using OCR (Optical Character Recognition) to read out the counter from the uploaded images in order to get reliable electricity usage figures proofed impossible due to these issues. In a bespoke software solution a better way to collect electric meter data must be found, ideally something similar to the barcode scanner, where the user has to explicitly focus the camera on the electric meter counter so that the application can read out the number and store the input numerically in the data. This is technically not trivial though [3, 4].

Finally, it would be useful if the barcode scanner implemented in the application could be automatically linked to an extensive database of barcodes, that would allow for finding the product information (incl. whether the product is eco-friendly, fair-trade, the price, etc.) linked to the barcode number and add this product information to the data directly. Unfortunately free and comprehensive barcode databases are currently not available, though there are first attempts to build such databases (e.g. http://product-open-data.com). There is also the problem of different types of barcode formats (e.g. EAN 13, UPC, etc.) and usually databases provide comprehensive data on one or the other, rarely on various types of barcodes. But, there are barcode database providers who offer APIs that would facilitate an automatic linking to these databases, even though these services are not free (e.g. https://www.barcodelookup.com/api).

S1.1. Bespoke Smartphone Application

Given the technical limitation of using EpiCollect 5 for the envisioned research discussed in the main article, work is ongoing on developing a bespoke smartphone application. However, for a major study it may be sensible to consider outsourcing the development of a professional smartphone application to a software company specialized on developing mobile phone applications, since developing a complex, multi–functional, high quality mobile phone application is not trivial. At the very least it will be necessary to considerably increase the resources for developing a bespoke application. In this section a brief description of our work on a bespoke smartphone application will be provided. Further technical details can be obtained from [5].

The bespoke application dataUp, that we have been working on, is a hybrid application, i.e. a hybrid between a native application and a web application. Hybrid applications have the advantage that they allow cross-platform 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) 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 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 format, 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 visualization 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 visualization of the data in the social monitoring intervention group it was decided to use JqPlot, one of the most well known Java Script visualization libraries, using HTML5 Canvas.

Messaging for the behavioral 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 a hybrid 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 barcodes and GPS recording. Moreover the application required further debugging and testing before being 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 continued.

S2. Data Collections

Study participants had to enter data through the application on a daily basis and upload the data in the evening. Study participants were asked to enter the start and destination GPS coordinates, barcode scans of products they bought, moreover they were asked what transport mode (multiple choice question) they chose (incl. no, if they stayed at home) and why (single choice question), what food they have eaten throughout the day (multiple choice question), what electric devices they used (multiple choice question), what electric devices were on all day (multiple choice question), what non-grocery products they have bought (multiple choice question) and what waste they produced (multiple choice question). They were also asked to take an image of their electric meter counter. Participants could skip questions requesting GPS coordinates, barcode scans and electric meter images, if they were uncomfortable with sharing this type of data or if they simply did not travel, or did not buy anything on a given day or if they had no access to their electric meter or were traveling. The full questionnaire implemented in the smartphone application is given below.

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 one 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 than 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 averages 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. Moreover, using EpiCollect 5 comes with limitations and restrictions, e.g. EpiCollect 5 does not allow for repeated collection of 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 behavioral targeting group were individualized, based on the users environmental performance. For instance, if the CO2 emissions were particularly high due to the 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

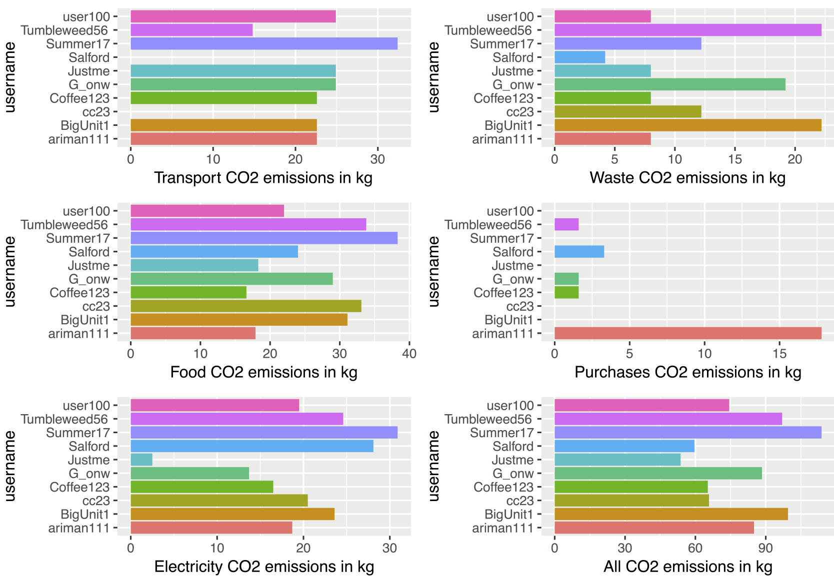

Fig. 1. Exemplary image that study participants in the social monitoring treatment group would receive

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 traveling 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 due 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."

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. Full Results

The results are discussed here only to show what analyses could be done and what insights could potentially be reached if data would be collected on larger scale using the suggested approach. No attempts will be made to make any (general) conclusions about environmental behavior from this pilot study data. Not only does the extremely small, unrepresentative sample not allow for any robust estimations or generalizations, but the two weeks of data collection are too short to measure any consistent behavioral changes due to field-experimental interventions. Indeed, in the evaluation survey after the data collection, some study participants explicitly said that they were not able to make short-term changes after noting their high environmental footprint scores during the field-experimental intervention week, because flights were already booked long time ago, car trips already planned and certain foods already bought. Finally, the lack of a control group limits the analyses that can be performed and the results that can be obtained.

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 "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.

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) raise 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.

The actual environmental behavior 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

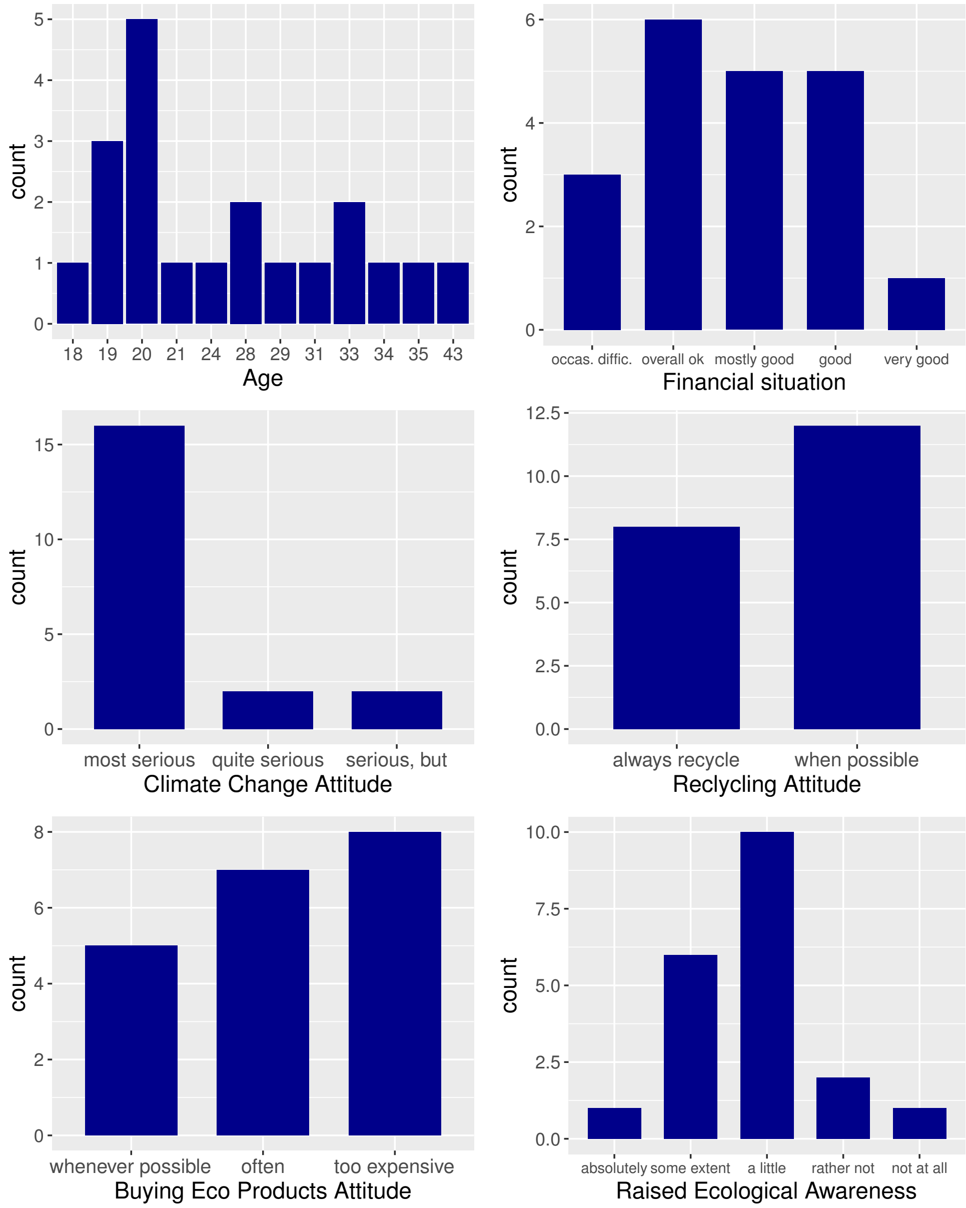

Fig. 2. Bar charts for various variables from the initial and final survey. Age ranged between 18 and 43, with the mean of 25.7 and standard deviation of 7.23. Financial situation ranged between 1 “very difficult” and 7 “very good”. The median of this variable was 5, “mostly good”. Climate change attitudes 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

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, i.e. data collection took place in June and students were on term break and were more likely to travel.

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. Indeed, 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

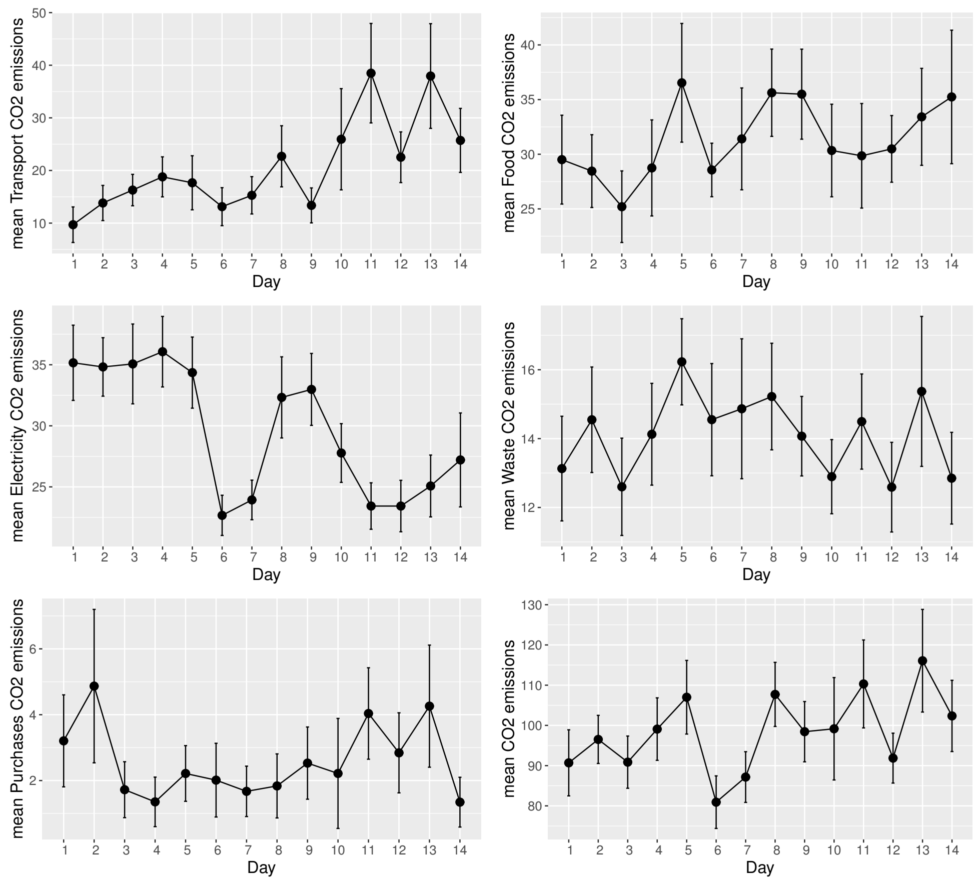

Fig. 4. Mean CO2 emissions, incl. standard error bars over the 14 days of data collection

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 there are serious issues with retrieving the data from the images of the electric meter counter as discussed in section S1. 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 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.

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

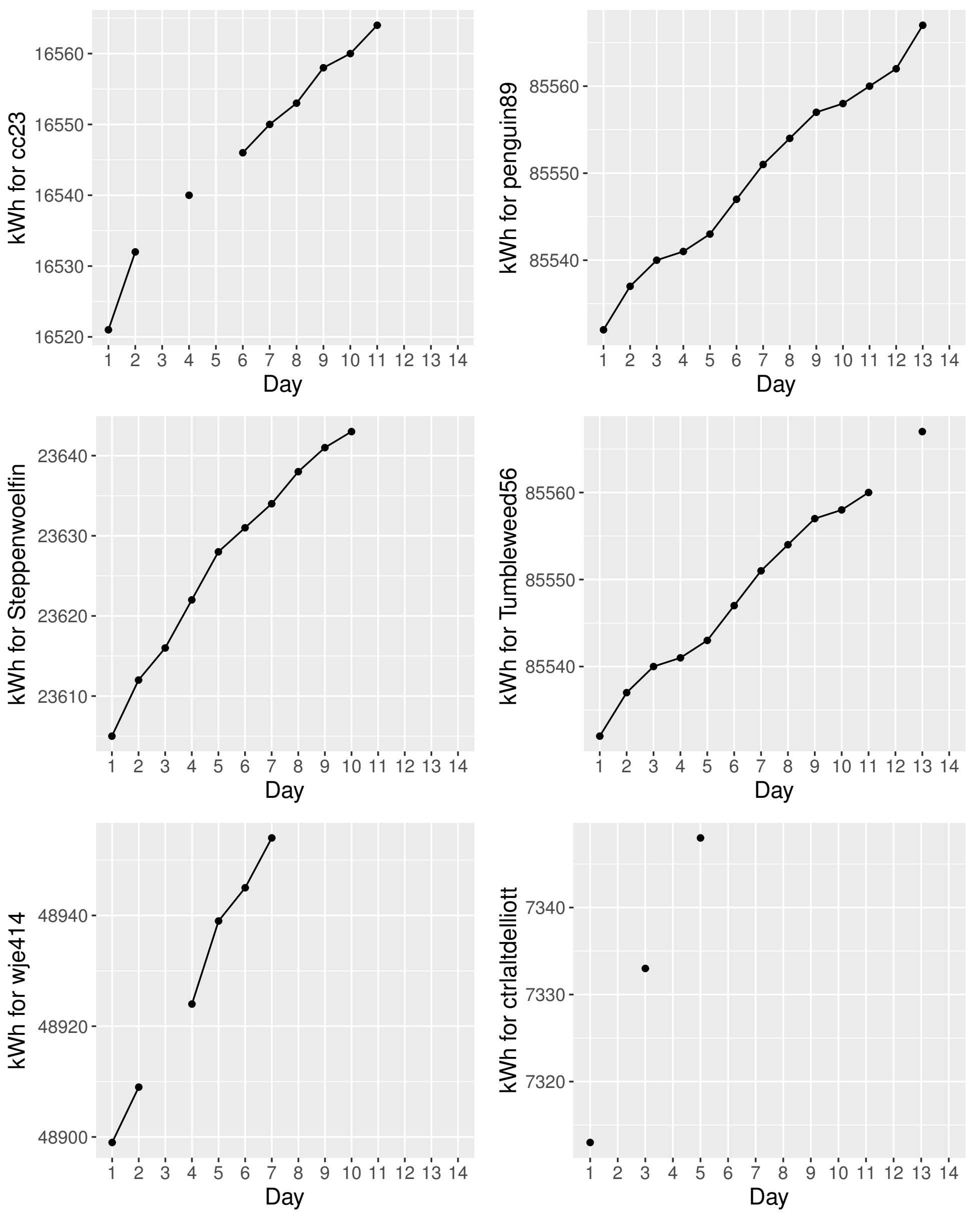

Fig. 5. Electricity usages based on electric meter data for six study participants over 14 days of data collection

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. But, barcode scan data gives, if available and decodable, much more detailed information that could be further exploited (e.g. precise item, quantity, price). Given this information richness, barcode scans should be preferred over survey responses in an actual study, though that would require participants' collaboration and commitment.

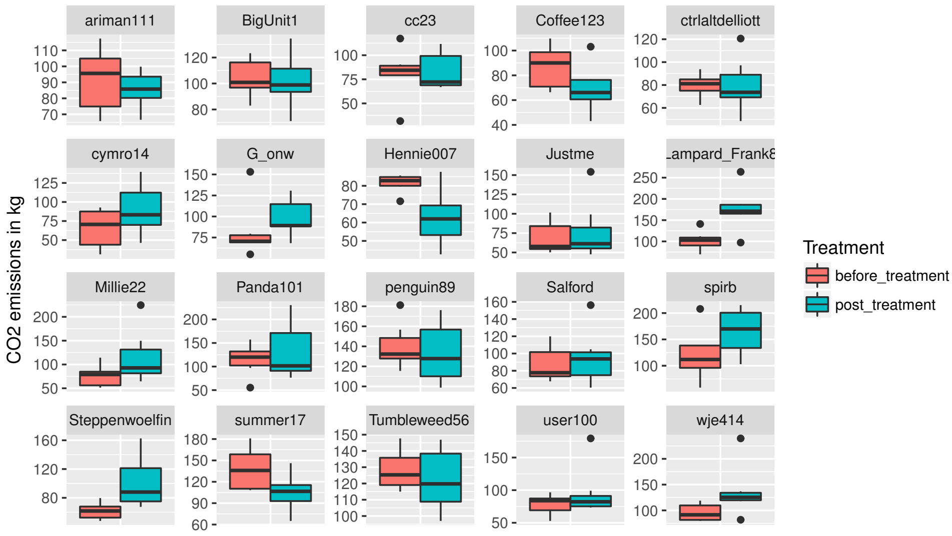

Fig. 6. Box plots for the 20 study participants showing their overall CO2 emissions before and after the field-experimental treatment.

S3.2. Treatment Effects

It is of particular interest, at least in an actual study, to estimate the treatment effects and to compare various field-experimental interventions. Figure 6 shows the overall CO2 emissions in the first and in the second week for the 20 study participants. It shows no coherent picture, in some cases the overall 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). This is not surprising, since, as already mentioned earlier, people need more time (min. 18 days!) to adjust their behavior [6]. One week of intervention is not sufficient to measure consistent behavioral change. People cannot be expected to change their traveling plans (e.g. cancel a booked flight) suddenly or throw away food they have bought earlier after an intervention. This is also evident if we conduct a paired t-test comparing the average overall CO2 emissions from the first and second week. The paired t-test shows that there is no significant treatment effect ($t = -2.12$, $p = 0.05$, 95%-CI: [-24.46, -0.13]), in fact, the mean difference (-12.29) even points in the opposite direction, that is, the aggregate mean CO2 emissions in the second week (after intervention) (106.49, sd = 30.36) was even somewhat higher than in the first week (94.19, sd = 23.46). This was partly 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 book their flights some considerable time in advance, it is safe to assume that the sudden rise of flights in the second week was not due to but rather despite the field-experimental treatment.

We also decomposed the overall CO2 emissions measure into the various environmental dimensions. Figure 7 shows CO2 emissions descriptives before and after treatment for each study participant for the transport dimension. The results resemble quite strongly the ones obtained for the overall CO2 emissions: 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.

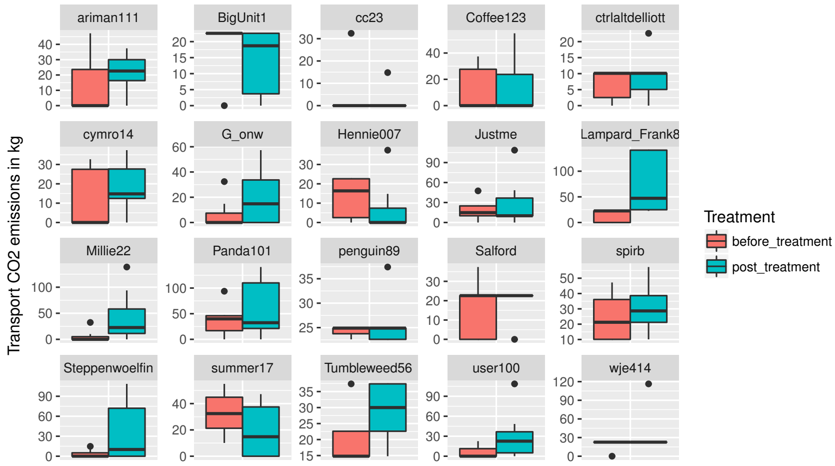

Fig. 7. Boxplots for the 20 study participants showing their transport CO2 emissions before and after the field-experimental treatment.

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 above, this is due to the fact that many study participants were flying in the second week.

For the food dimension, we can see the CO2 emissions descriptives before and after treatment for each study participant in Figure 8. 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 difference between the mean food CO2 emissions before and after the field-experimental intervention (t = -2.32, p = 0.03, CI: [-7.91, -0.41]) and the mean difference of -4.16 points clearly in the opposite direction of the expected treatment effect. 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 9 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 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 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

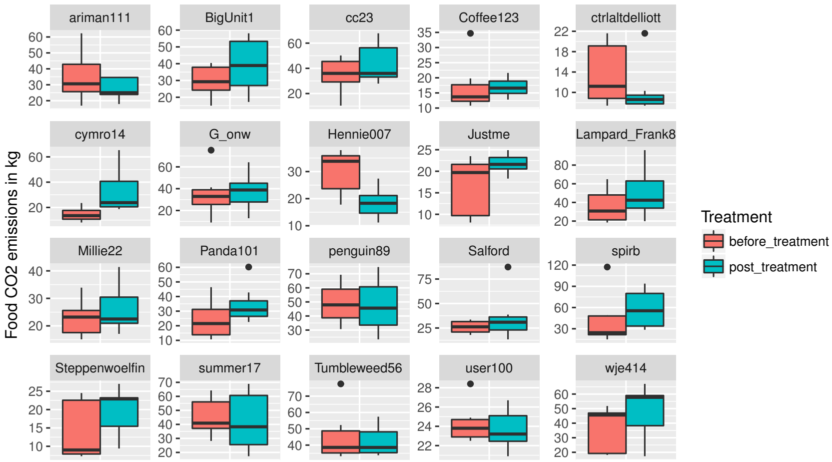

Fig. 8. Boxplots for the 20 study participants showing their food CO2 emissions before and after the field-experimental treatment.

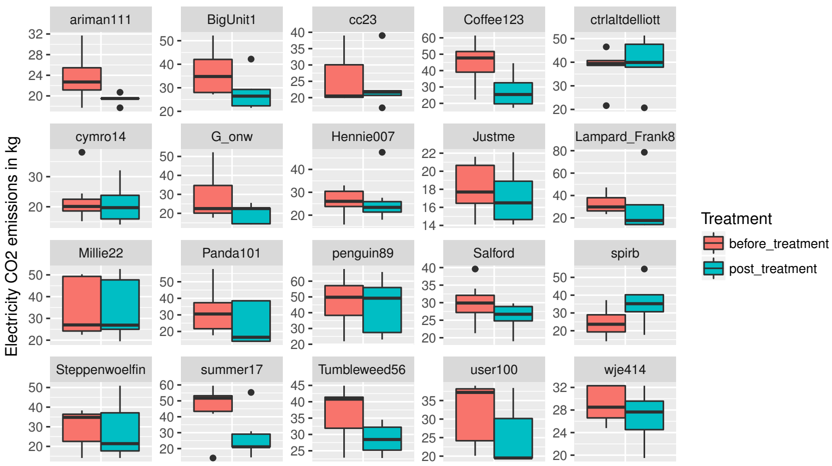

Fig. 9. Boxplots for the 20 study participants showing their electricity CO2 emissions before and after the field-experimental treatment.

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 10 and 11 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 also be 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

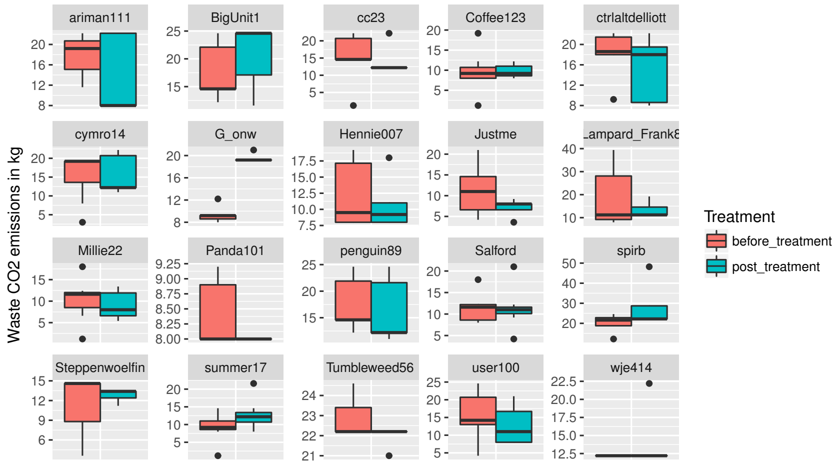

Fig. 10. Boxplots for the 20 study participants showing their waste CO2 emissions before and after the field-experimental treatment.

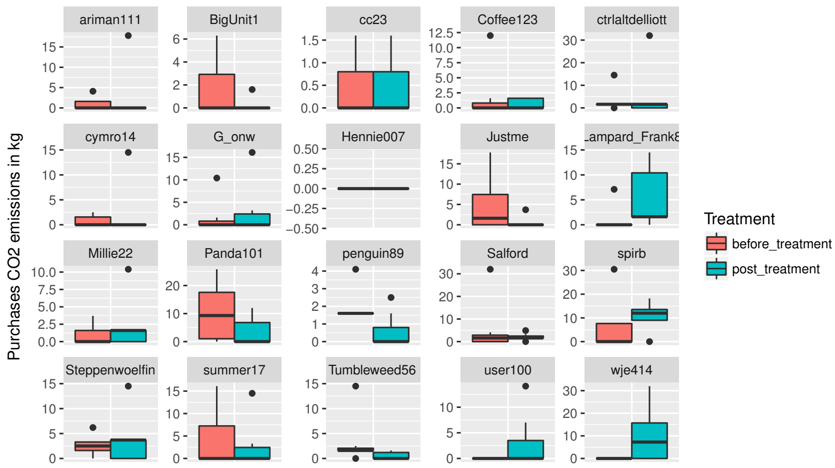

Fig. 11. Boxplots for the 20 study participants showing their purchases CO2 emissions before and after the field-experimental treatment.

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).

Overall we can see that it's not only the transport dimension that is inconsistent with our assumptions.

The pilot study does however not really allow us to draw any conclusions, but it would be interesting to investigate, whether a treatment may even have a counter-productive effect, at least in some dimension or at least at the start of the intervention. People seem to be particularly resistant to interventions if it comes to their food [7]. Only in the electricity dimension we can see a significant positive treatment effect. People find it probably easier to save some energy in response to a field-experimental treatment than to change their eating habits or their traveling plans. Thus, the pilot study would suggest that the two treatments did not change study participants' behavior in positive (greater sustainability) way, except for in the electricity dimension. However, as mentioned earlier, this was rather to be expected because behavioral adjustments takes usually more time [6]. Also, due to financial limitations which prohibited the recruitment of further study participants, no control group is available, which limits the analyses, since only a comparison against the past performance over the first week is feasible. The analyses also show considerable variation, incl. outliers in the CO2 emissions, which makes a paired t-test on individuals' average CO2 emissions rather problematic.

S3.3. Repeated Measures ANOVA

The above mentioned problem can be avoided to some extent in a repeated measures ANOVA that allows to compare the two treatment effects, using second week data, accounting for autocorrelation and random effects. The analyses result in a reasonable random effect model (LogLikelihood (LL): -661.11, AIC: 1336.23, BIC: 1356.19) that fits the data indeed better (L-Ratio Test: 16.29, $p < 0.01$) than a fixed model (LL: -669.26, AIC: 1348.52, BIC: 1362.78). Calculating Pseudo-R-squares, which are relative fit measures comparing various models, based on McFadden (0.02), Cox/Snell (0.17) and Nagelkerke (0.17) for the model vs. a null models with neither fixed nor random effect, shows moreover that the model fits the data better than a null model (LL-difference: -12.30, Chi-Square: 24.61, $p < 0.01$). However, the model is only marginally better than a null model that includes random effect and the improvement in the model fit is not significant (LL-diff: -2.36, Chi-Square: 4.72, $p = 0.32$; McFadden Pseudo R-Square: 0.004, Cox/Snell Pseudo-R-Square: 0.35, Nagelkerke Pseudo R-Square: 0.35). This suggests, that at least with the obtained data we cannot necessary conclude that accounting for the treatments improves our ability to predict the overall CO2 emissions. This is likely due to the very small data set. Nevertheless, when actually comparing the two treatments, the pilot study seems to suggest that the social monitoring treatment had a somewhat greater positive (in terms of reducing CO2 emissions) effect on the overall environmental behavior in comparison to the behavioral targeting treatment (see Figure). This result shows that it could be worth investigating further the effect of different types of treatments in a full study, including a control-group and potentially other treatments.

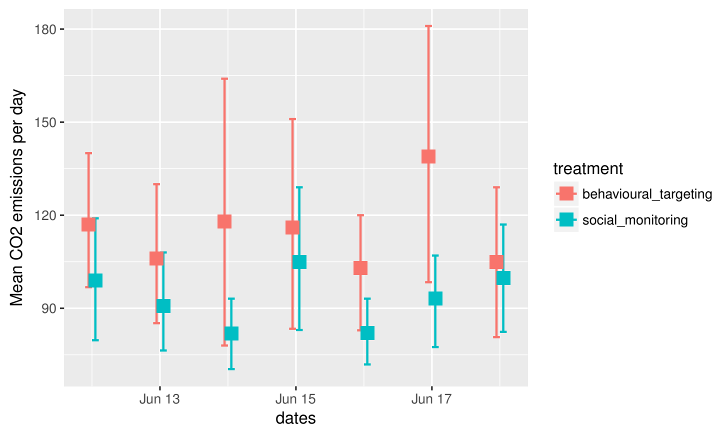

Fig. 12. Interaction plot shows the natural mean of each treatment*date combination along with the confidence interval of each mean with percentile method. Least Square Means for CO2 emissions in the behavioral targeting group were estimated to be 118.02 (se: 8.61, 95%-CI: [97.12, 138.93]), for social monitoring 93.19 (se: 8.52, 95%-CI: [72.42, 113.96]).

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 (LL) 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															
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		

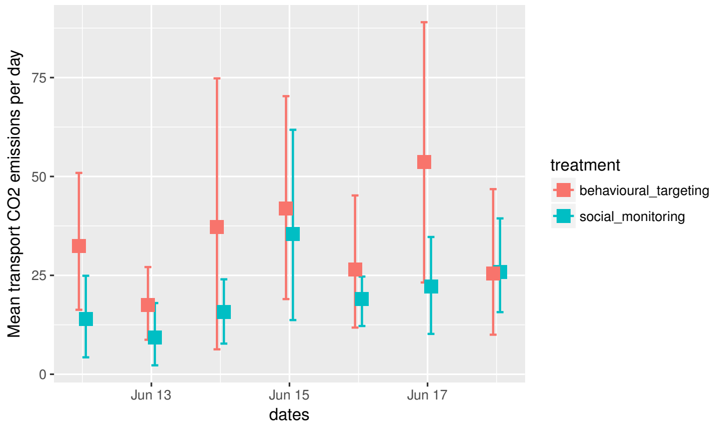

Fig. 13. Interaction plot for transport dimension shows the natural mean of each treatment*date combination along with the confidence interval of each mean with percentile method. Least-Square means for CO2 emissions in the behavioral targeting group were estimated to be 34.28 (se: 5.22, CI: [21.60, 46.96]), for social monitoring 20.19 (se: 5.12, CI: [7.70, 32.67]).

The Repeated Measures ANOVA results for the different environmental behavior 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 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 the random effect model is not very strong. Moreover, 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 behavioral targeting (see Figure 13).

The model fit of the random effect model for the food dimension CO2 emissions is on the other hand 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. But, it is less clear in this case, which of the two treatments has the strongest effect on behavioral change in terms of reducing CO2 emissions (see Figure 14).

The paired t-tests discussed above suggested that if the treatments had any effect on environmental behavior 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 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 15 and the Least Square Means analyses show, the social monitoring treatment has a stronger and more

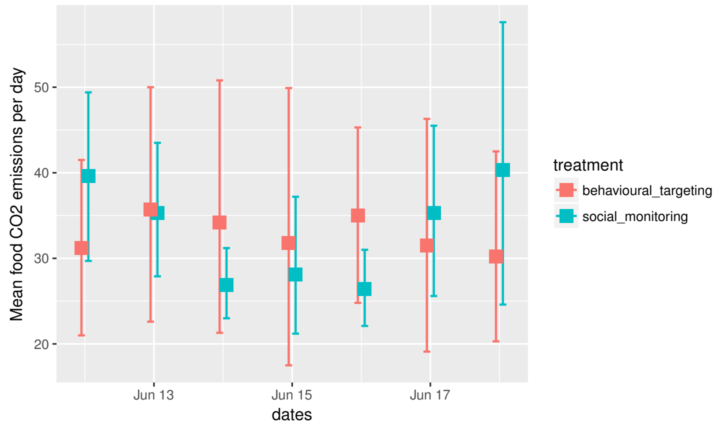

Fig. 14. Interaction plot for food dimension shows the natural mean of each treatment*date combination along with the confidence interval of each mean with percentile method. Least-Square means for CO2 emissions in the behavioral targeting group were estimated to be 35.08 (se: 4.56, CI: [24.01, 46.15]), for social monitoring 34.44 (se: 4.54, CI: [23.38, 45.50]).

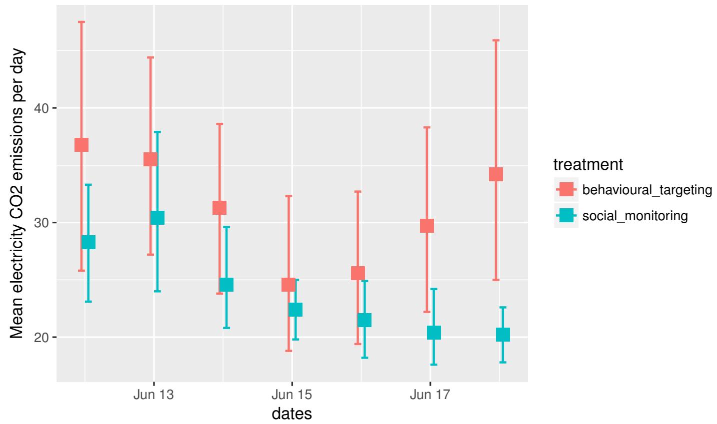

Fig. 15. Interaction plot for electricity dimension shows the natural mean of each treatment*date combination along with the confidence interval of each mean with percentile method. Least-Square means for CO2 emissions in the behavioral targeting group were estimated to be 31.78 (se: 1.94, CI: [27.06, 36.50]), for social monitoring 24.18 (se: 1.92, CI: [19.50, 28.86]).

consistent effect on behavioral change in terms of electricity-based CO2 emissions reductions comparing to behavioral 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, similarly to the food dimension, it is less clear in this case, which of the two treatments has the strongest effect on behavioral change in terms of reducing CO2 emissions (see Figure 16). 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

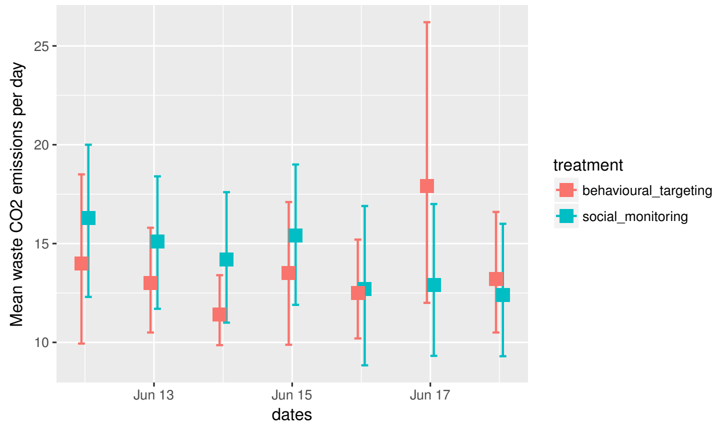

Fig. 16. Interaction plot for waste dimension shows the natural mean of each treatment*date combination along with the confidence interval of each mean with percentile method. Least-Square means for CO2 emissions in the behavioral targeting group were estimated to be 14.61 (se: 1.77, CI: [10.33, 18.89]), for social monitoring 14.54 (se: 1.76, CI: [10.25, 18.83]).

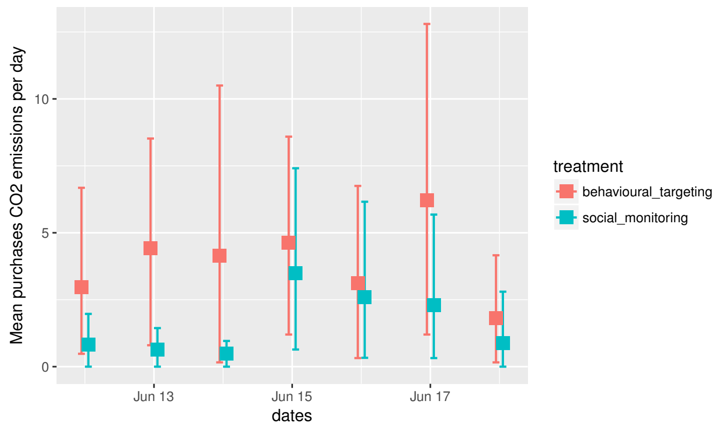

Fig. 17. Interaction plot for purchases dimension shows the natural mean of each treatment*date combination along with the confidence interval of each mean with percentile method. Least-Square means for CO2 emissions in the behavioral targeting group were estimated to be 4.06 (se: 0.79, CI: [2.13, 5.99]), for social monitoring 1.63 (se: 0.77, CI: [0, 3.52]).

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, though 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 the model is not very strong. The treatments thus seem to have essentially no predictive power when it comes to purchases CO2 emissions, which seems generally not to follow a predictive pattern potentially due to insufficient data. Nevertheless, if any treatment has any effect, then it seems to be the social monitoring treatment rather than the behavioral targeting (see Figure 17).

Overall, the repeated measures ANOVA analyses for the different environment dimensions confirm among others that the strongest effect is measurable in the electricity dimension, with the social monitoring treatment having a stronger and more consistent effect comparing to the bahavioral targeting treatment.

S3.4. Gaussian Processes Choice Models

The data collected also allows for building and estimating choice models, either traditional discrete choice models [8] or if one would like to account for potential non-linearities in the decision making and in general permit a Machine Learning algorithm to find the best utility function to describe the choice data, then the Processes-based choice models suggested recently by Mann et al. [9, 10] are quite useful. Gaussian process choice models are essentially a non-parametric extension of the conditional logit model, using Gaussian process priors. The conditional logit model is used in social science for inferring interaction effects between personal features and choice characteristics from observations of individual multinomial choices, such as where to live or what transport mode to use. The classic, parametric model presupposes a latent utility function that is a linear combination of choice characteristics and their interactions with personal features. The Gaussian process choice models allow for non-linear utility functions while controlling for model complexity [9].

Gaussian process choice models were estimated for transport mode choices. Initial survey data provided individual characteristics data (i.e. age, financial situation, climate change attitude). While data collected through the smartphone application allowed for the creation of a dataset of travel choices characteristics (e.g. CO2 emissions), accounting for distances of individual travels, estimated from the GPS recordings, to which other data on travel choices from other sources can be added (e.g. average speed, average cost for a respective travel mode). Finally, using the data from the smartphone application, that indicates who used what transport mode when, why and to travel where etc. the individual characteristic data is linked to the travel characteristic choices. Making use of Mann et al. [9, 10] approach, we estimated the utility functions to understand why study participants have chosen certain transport modes in a given situation. Keeping in mind the limitation of the data, the obtained model results nevertheless show the potential of this approach to gain insights from an actual study.

Figure 18 shows that distance plays an important role when it comes to picking the travel mode and hence when it comes to CO2 emissions. Participants prefer for small distances low CO2 emission transport modes, but with increasing distance transport modes with higher CO2 emissions are preferred and in fact are becoming increasingly without any alternative. This relation seems to interact to some extent with the financial situation. The transport mode choices of less affluent participants seem to be more limited, the utility bands in the plot are much more narrow and focussed. But the main two positive utility areas for lower and higher distances are the same for the well-off and less well-off. Age seems to have a rather minor effect on its own, in particularly it does not play a major role for travels within low distances and for middle distances the older seem to have a slightly higher preference for high-CO2 emission transport modes compared to younger participants, who might have less access to these (i.e. car ownership). The choice is furthermore correlated with climate change attitudes. For near distances those, who are more concerned about the climate change, are more likely to prefer low-CO2 emission transport modes, while for far distances those who are less concerned are having a much clearer preference for high-CO2 emission transport modes. Generally, those who are least concerned about climate change are more likely to prefer transport modes with high CO2 emissions even for small distance travels. Besides CO2 emissions, participants seem to choose transport modes based on how much independence these transport modes allow for, with a clear preference for transport modes that allow for the greatest independence such as cars, but also bikes and walks for short distances, while transport modes that are low

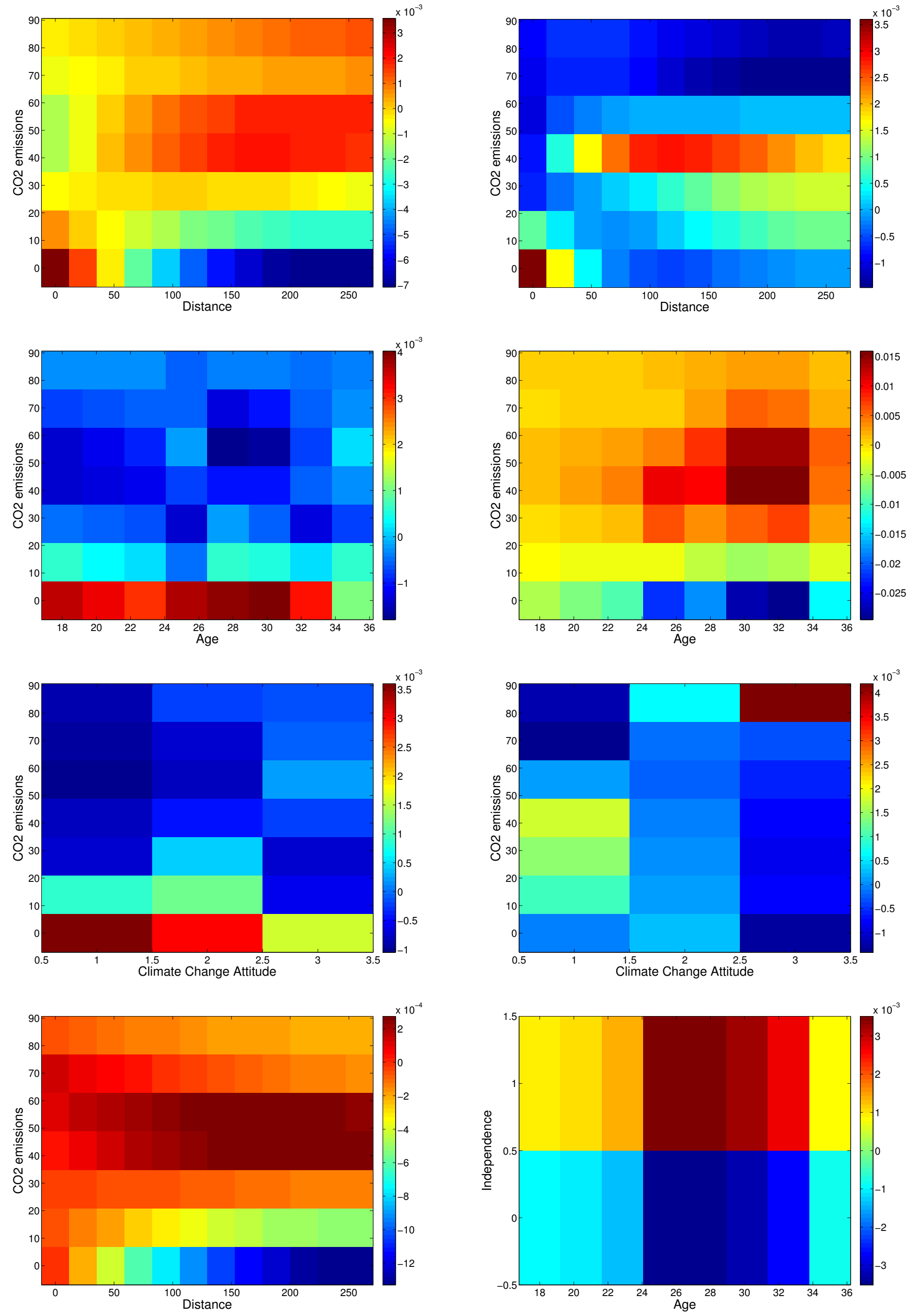

Fig. 18. Heat plots displaying the utility function for transport modes based on choice characteristics CO2 emissions and independence. The color bar shows the utility scale, with redder colors indicating a positive utility and bluer colors none or even negative utility. The upper two panels show transport mode preferences with respect to travel distance for younger and financially well (left) or less affluent (right) participants (a similar pattern emerges for older participants, see Supplementary Information S3.4). The two panels in the second row show the effect of age, holding financial situation constant, comparing near distance travels (left) and middle distance travels (right). The two panels in the third row display climate change attitude effects, comparing near distance travels (left) and far distance travels (right). The bottom left panel shows the effect of distance, if climate change attitude is held constant at "least concerned" level for young and well off participants. The bottom right panel shows utility function for transport modes based on independence with respect to age, for affluent participants.

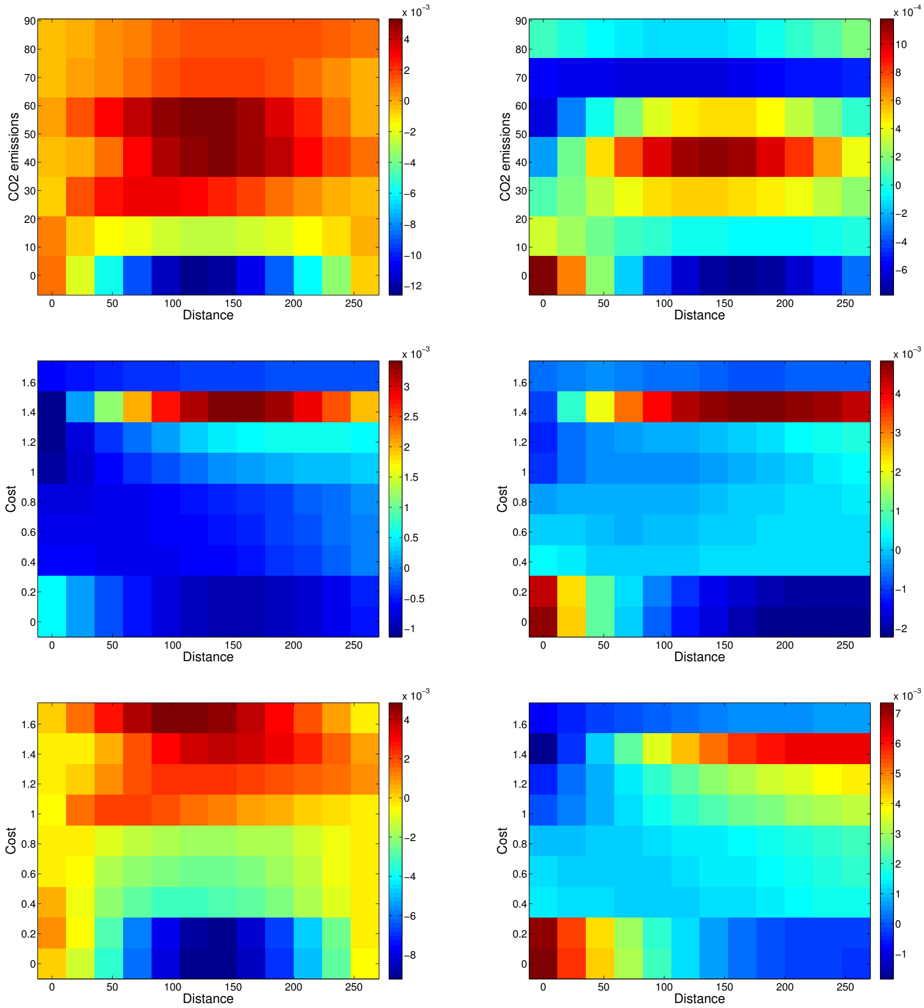

Fig. 19. Heat plots displaying the utility function for transport modes based on CO2 (upper two panels) and costs as choice characteristics. The color bar shows the utility scale, with redder colors indicating a positive utility and bluer colors none or even negative utility. The upper two panels show transport mode preferences with respect to travel distance for older and financially well (left) or less affluent (right) participants. The two panels in the second row show the effect of distance for older but less affluent participants (left) and younger less affluent 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).

on independence such as buses and trains are rather disliked. Moreover, we see that older participants seem to have a slightly higher preference for independent transport modes comparing to young participants, who again might not have access to these (e.g. car ownership). We can also 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 less affluent (see Figure 19, upper two panels) as they did for younger participants.

When it comes to travel costs as a choice characteristic, the results suggest that study participant do indeed make a transport choice taking costs into account. As Figure 19 shows, older and younger less affluent participants have a similar preference for a certain transport mode of a certain cost once the distance prohibits taking lower cost transport modes, for which the younger have a slightly stronger preference. However this might be an artifact of lack of data, since the less affluent 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 long distances 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 less affluent 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.

S3.5. Interactions between various environmental dimensions

Finally, of interest is also to investigate how the various environmental behavior dimensions interact, specifically whether we can find evidence for the so-called moral credential effect or self-licensing effect [11, 12], where a person who has chosen an environmentally friendly behavior in one context, may feel morally entitled to behave in less environmentally friendly fashion in another. On the other hand a person may decide to compensate for an environmentally damaging behavior though a particularly environmentally friendly behavior in another dimension. Looking bivariately at the correlations between the various environmental dimensions suggests however only weak relations between the dimensions and mostly positive, hence suggesting that study participants with high CO2 emissions on one dimension tend to be higher on the others (see upper Figure 20). Weak evidence for compensation behavior arises only in the relation between transport and electricity, where we see a negative correlation, though this might be an artifact, i.e. on days when study participants travel they automatically tend to use fewer electronic devices and hence the electricity CO2 emissions are lower.

When considering only week two data (see bottom Figure 20) the correlations are a bit stronger, but still weak and mostly insignificant (e.g. the negative correlation between Transport CO2 and Electricity CO2 in the second week is -0.198, for the whole study period it is -0.146). Moreover, looking at correlations taking into account lags, which would allow to investigate moral credential effects, there is a very weak and insignificant negative correlation between lag Electricity CO2 and Transport CO2 ($r = -0.051$), which could potentially hint at a moral credential effect, where study participants who saved electricity (and hence CO2 emissions) the previous day felt potentially a bit more entitled to use less environmentally friendly transport modes the next day. But the pilot study data is insufficient to provide valid evidence for this assumption. Given these very weak bivariate interactions further model-based investigations were not conducted, but in a full-scale study it would be still interesting to take a closer look at the interactions of the various environmental behavior dimensions. The small data set and other pilot study limitations may have inhibited measuring clearer relations.

The pilot study results presented here mostly show that potentially interesting insights could be gained from conducting such a study with an improved design, on a larger scale and over a longer period. Some results were non-conclusive due to various limitations and it remains to be seen if better data that would allow for better statistical analyses (e.g. mixed effect models, difference-in-difference analyses etc.) could produce clearer outcomes. Furthermore, 17 out of 20 study participants suggested that the participation in the study increased their environmental awareness a little (10), to some extent (6) or absolutely (1). Of course this does not translate automatically in behavioral change, but there may be some potential for it. Generally, it is also quite indicative that although the vast majority (16 out of 20) of the study participants agreed that climate change is one of the most serious problems that humanity is facing today' (see section S3.1), this concern is not necessarily mirrored in their everyday environmental behavior; cars and taxis are used on a regular basis even for shorter distances, beef or lamb is consumed, even though it has an extremely high carbon footprint etc.. It shows how abstract climate change concerns remains for most people when it comes to their day-to-day life. Attitudes do not necessarily manifest themselves in behavior.

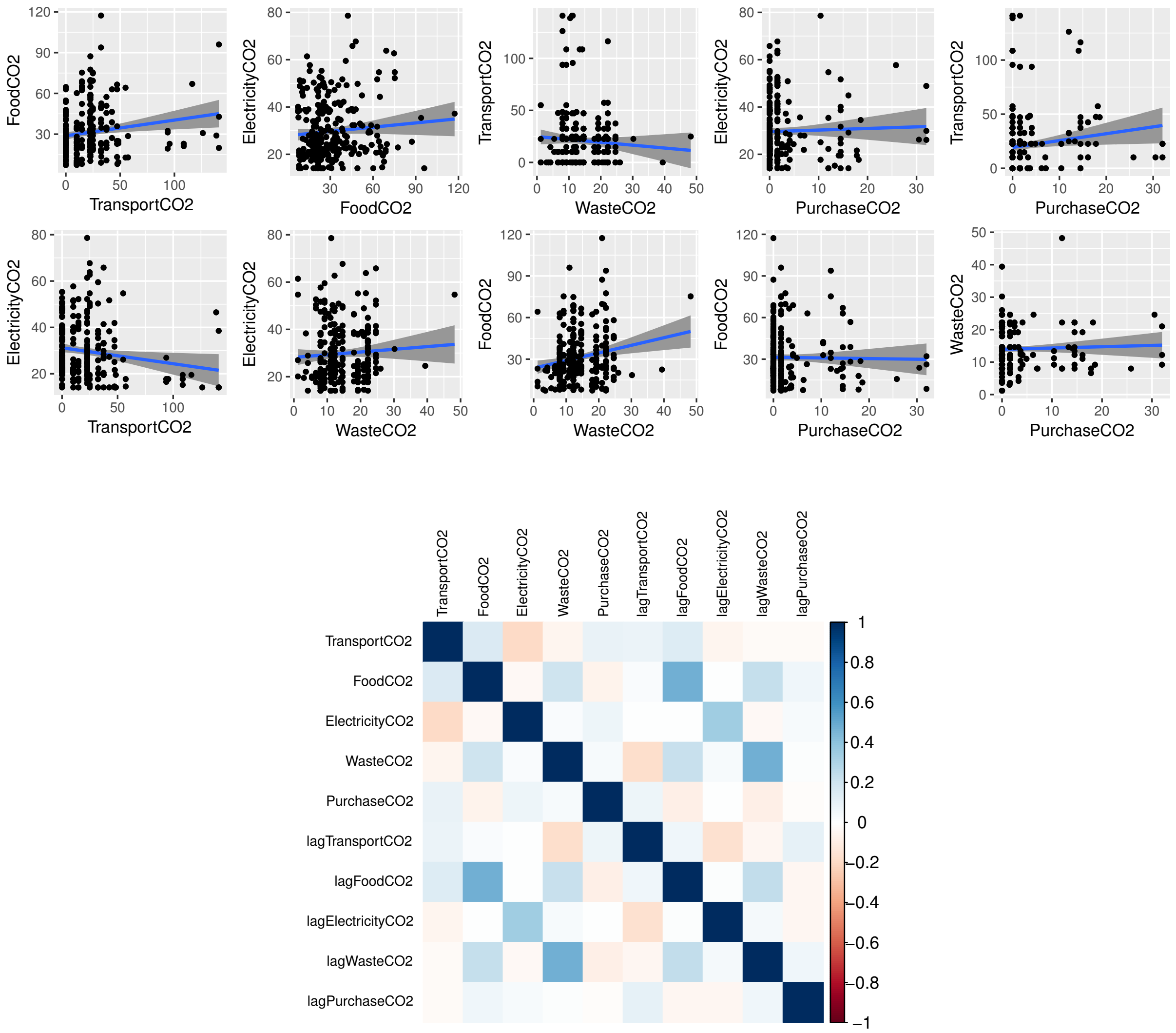

Fig. 20. The upper panels are scatterplots with a fitted linear regression line and 95%-Confidence Interval areas around it to display the bivariate relation between the various environmental dimensions. The bottom figure is a correlation matrix plot for second week data that displays the direction and strengths of correlations between the various environmental dimensions in colors, including one-day lags.

References

[1] D.M. Aanensen, D.M. Huntley, E.J. Feil, F. al-Own and B.G. Spratt, EpiCollect: Linking Smartphones to Web Applications for Epidemiology, Ecology and Community Data Collection, *PLoS ONE*, **4** (2009), e6968. doi:10.1371/journal.pone.0006968.

[2] D.M. Aanensen, D.M. Huntley, M. Menegazzo, C.I. Powell and B.G. Spratt, EpiCollect+: linking smartphones to web applications for complex data collection projects, *F1000Res*, **3** (2014), 199. doi:10.12688/f1000research.4702.1.

[3] M. Petter, V. Fragoso, M. Turk and C. Baur, Automatic text detection from mobile augmented reality translation, *2011 IEEE International Conference on Computer Vision Workshop (ICCV Workshops)*. doi:10.1109/ICCVW.2011.6130221.

[4] T.S. Gunawan, A. Mutholib and M. Kartiwi, Performance Evaluation of Automatic Number Plate Recognition on Android Smartphone Platform, *International Journal of Electrical and Computer Engineering*, **7** (2017), 1973–1982. doi:10.11591/ijece.v7i3.pp1973-1982.

[5] A. Gregoriou *Application for Environmental behavior 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.

[6] P. Lally, C.H.M. van Jaarsveld, H.W.W. Potts and J. Wardle, How are habits formed: Modelling habit formation in the real world, *European Journal of Social Psychology*, **40** (2010), 998–1009. doi:10.1002/ejsp.674.

[7] B. Bastian, S. Loughnan and N. Haslam, Don't Mind Meat? The Denial of Mind to Animals Used for Human Consumption, *Personality and Social Psychology Bulletin*, **38** (2012), 247–256. doi:10.1177/0146167211424291.

[8] M.E. Ben-Akiva and S.R. Lerman, *Discrete Choice Analysis: Theory and Application to Travel Demand*, MIT Press, 1985. ISBN: 9780262022170.

[9] R.P. Mann, V. Spaiser, L. Hedmann and D.J.T. Sumpter, *Choice Modelling with Gaussian Processes in the Social Sciences: A Case Study of Neighbourhod Choice in Stockholm*, SSRN, (2016). https://papers.ssrn.com/sol3/papers.cfm?abstract_id=2869215

[10] R.P. Mann, *Gaussian process conditional logit model*, MATLAB code on GitHub, 2016, https://github.com/richardpmann/gp_conditional_logit.

[11] A.C. Merritt, D.A. Effron and B. Monin, Moral Self-Licensing: When Being Good Frees Us to Be Bad, *Social and Personality Psychology Compass*, **4** (2010), 344–358. doi:10.1111/j.1751-9004.2010.00263.x.

[12] J. Jordan, E. Mullen and J.K. Murnighan, Striving for the Moral Self: The Effects of Recalling Past Moral Actions on Future Moral Behavior, *Personality and Social Psychology Bulletin*, **37** (2011), 701–713. doi:10.1177/0146167211400208.