Data Science x (2017) 1–14
IOS Press

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

Abstract. Ecological sustainability is the defining challenge of our time. This paper suggests a new methodological approach that could help to investigate how environmental behaviour (transport behaviour, energy consumption, food consumption, goods consumption, wasting) dilemmas can be overcome on an individual level in real life by using smartphones to collect daily behavioural data in a field-experimental setup. Presented are results from a pilot study to discuss the feasibility and potential of this approach. After the first week of data collection without any intervention, the 20 study participants were randomly assigned to two field-experimental treatment groups. In the social monitoring experimental group study participants monitored each other's environmental performance and in the behavioural targeting treatment group study participants' past behaviour was analysed to provide individualised tips how they can improve their environmental performance. The paper shows how the collected data could be used to better understand how people make environmental behavioural decisions, what can positively affect this decision–making in real life and how the different environmental behaviours (e.g. transport behaviour vs. energy consumption) interact with each other in real life.

Keywords: field-experiment, smartphone data, environmental behaviour, choice modelling, treatment effects

1. Introduction

In 2015 the United Nations implemented its new Sustainable Development 15-years agenda (https://sustainabledevelopment.un.org). Several of the 17 Global Sustainable Goals are dedicated to preserving the environment (e.g. mitigating climate change, protecting marine systems, protecting forest systems etc.). The challenge that nations worldwide now face is how to make the transition towards a sustainable society. At the core of this challenge lies the social dilemma problem: a preserved environment is a common good of benefit to everyone; to achieve sustainability, however, cooperation is required from the majority. But, cooperation comes at individual costs in the short term and this provokes noncooperative behaviour [1]. This paper suggests to study environmental behaviour in real life social dilemma situations

*Corresponding author.

2451-8484 © 2017 – IOS Press and the authors. All rights reserved

by exploiting smartphone technology to collect new types of "living laboratory" [2] data. The novelty is thus to fuse "big" data (multiple format data collected via smartphones) with a theory-based field-experimental approach to study human behaviour in real life.

Big data is widely regarded as a rich data source for (environmental) human behaviour [3, 4]. Typically, consumer behaviour is the focus of environmental behaviour studies making use of big data such as retailers' loyalty cards data [5] or smart meter data [6]. However, such data is limited. For instance, Hornibrook et al. [5] could not explain why the introduction of carbon emission labelling on supermarket products did not have any impact on customers' purchase choices; the loyalty card data was not sufficient to answer this question and the researchers had to conduct focus groups to get insight into possible reasons for the lack of impact. Big data is typically purely observational, not generated for scientific purposes, useful to answer certain exploratory questions, but problematic where specific (causal) mechanisms are of interest.

Experiments on the other hand allow to explore cooperation mechanisms and showed for instance that public goods can be produced only in the presence of repeated interactions, which facilitate reciprocation, reputation effects and punishments or relatedness [7]. But, studies have also shown that the correspondence between laboratory experimental and field-experimental results is often quite weak [8, 9], suggesting that we cannot necessary make conclusions about real life (social dilemma) behaviour from laboratory experiments. Consequently there is a lack of deeper theoretical understanding of how these dilemma mechanisms play out in real life [10] beyond non-generalizable case studies [1].

Consequently, the most recent methodological development aims to combine the big data approach with experimental design [11–13]. Mobile technologies can be ideal tools for such combined approaches [13–15]. Smartphones are for instance used to study people's daily lives, tracking social interactions, mobility routines, etc. [16]. However most of these studies do not implement an experimental design. The major originality of this pilot project lies in pushing the boundaries of what can be done with "living laboratory" data in order to better understand (environmental) social dilemma problems.

This research aims to contribute to a greater understanding of people's day-to-day environmental decision making and to investigate how environmental behaviour dilemmas can be overcome on an individual level in real life by combining the richness of "big" data with a field-experimental approach. Overcoming environmental behaviour dilemmas is essential to a successful transition to an ecologically sustainable society, as envisioned by the United Nations. Moreover, social dilemma problems such as the environmental dilemma problem, the volunteering dilemma or any cooperation dilemma are pervasive through social sciences [1] and a core issue in every society.

Studying multiple environmental behaviour dimensions (transport, food, energy, consumption, waste) simultaneously is crucial to understanding how they interact in people's decision making, e.g. when people decide to buy an ecological product to compensate for environmentally damaging travel behaviour. Including different environmental behaviour dimensions allows to investigate phenomena like the moral credential effect, where a person who has chosen an environmentally friendly behaviour in one context, may feel morally entitled to behave in less environmentally friendly fashion in another. The Randomised Controlled Trial (RCT) field-experimental design allows moreover causal inference in hypothesis testing. And with the time-series experimental design actual behavioural changes and the stability of these changes can be analysed.

In the pilot study two field-based interventions were tested to inspire cooperative, i.e., environmentally-friendly behaviour. (1) Behavioural targeting, originally an online advertisement practice where online users are presented with advertisement based on their past online behaviour [17], is based in the nudge theory [18]. The "nudge" is any aspect "of the choice architecture that alters people's behaviour in a predictable way without forbidding any options or significantly changing their economic incentives" (ibid.). In this study the nudges will be tailored messages sent to study participants' smartphones, proposing specific behavioural changes based on participants' past behaviour. (2) In the social monitoring treatment group individuals will mutually monitor each other's behaviour as captured by various environmental behaviour scores and visualised through the smartphone application. It is assumed that people, who are aware of their behaviour being monitored and who can compare their behaviour to peer behaviour, will tend to show socially desirable, i.e. environmentally friendly behaviour [19, 20]. This hypothesis is based in the social influence theory which investigates the effects of compliance, conformity and competition [21]. A control group was not included in the pilot study due to financial restriction that allowed to recruit and compensate only 20 study participants. Instead it was decided to issue the field-experimental treatment only in the second week of the study. The first week thus served as reference point for comparison and treatment effect estimation. In an actual study however a control group should be included. Moreover, other field-experimental treatments, e.g. reputation-based interventions or financial incentives could be implemented in an actual study.

This paper will discuss the results of pilot data analysis 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 behaviour 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. Ideally the study would run over several weeks to measure stable effects. One week of intervention does not give study participants enough time to adapt and potentially change their behaviour. 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.

2. Methodological Approach

2.1. 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 research data collection purposes [22, 23]. The platform allows to create 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, since 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 a great platform for research data collections. For an actual study on larger scale a bespoke software solution will be preferable, which would allow to implement certain features more directly. For instance, EpiCollect 5 is not necessarily designed for experimental research purposes; 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 visualising 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 individualised 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 allow to send messages directly and automatically (incl. reminders to upload data) to users through the app and to visualise environmental performance within the app (see further details in Supplementary Information S1).

Moreover, 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 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 [24, 25].

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 to find 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 databased (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).

2.2. Data Collection

For the pilot study 20 study participants were recruited, 13 participants were students (incl. two postgraduate students) and 7 had a professional background. All study participants had a higher educational background. The age of the study participants ranged between 18 and 43, with the mean of 25.7 and standard deviation of 7.23. 8 study participants were male, 12 female. Study participants were compensated for their participation with a £50 Amazon voucher. When recruiting study participants it was attempted to maintain the non-interference assumption, i.e. that the CO2 emissions of one experimental group are not affected by the treatment in the other experimental group [26] e.g. through a spill-over effect between study participants who are friends.

The data collection took place over two weeks in June 2017. Note, that this means that students had their term break around this time and were more likely to travel. Study participants had to enter data through the application on a daily basis and upload the data in the evening. If they failed to do so they received a reminder email. The daily data entry took between 5 and 12 minutes and could be distributed over the whole day. 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 (see Supplementary Information S2.1 for the full questionnaire). Participants were allowed to 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 answers were then translated into average CO2 emissions for the specified activity based on "How Bad are Bananas? The Carbon Footprint of Everything" by Mike Berners-Lee [27]. This allowed to calculate average CO2 emissions for each environmental behaviour dimension and overall.

At the start of the pilot study the study participants were moreover asked to complete an initial online survey, collecting some basic socio-demographic (i.e. age, gender, financial situation) and attitudinal data (e.g. attitude on climates change, recycling and buying ecological products). And the end of the pilot study they were asked to complete a final online survey, evaluating their experience as study participants (i.e. what they liked, what they did not like, what suggestions for improvement they have, whether they thought that participation in the study raised their environmental awareness, whether they thought the compensation was sufficient and whether they would participate in a similar study over a longer period of time) (see Supplementary Information S2.2. and S2.3 for further details).

In the first week of data collection no field-experimental treatment was issued. After the first week the 20 study participants were randomly assigned to one of the two field-experimental groups, each containing 10 study participants. Thus, in the second weak all study participants were subject to one of the two field-experimental treatments on a daily basis. In the behavioural targeting group they would receive individualised messages giving advice on how they could reduce their CO2 emissions, e.g. in the transport dimension by using a bus instead of a car. The advice given was based on the data entered on the previous day. In the social monitoring group study participants would receives messages visualising in a bar graph their own environmental performance from the previous day on the various dimensions as well as the environmental performance of the others in the group. This happened in an anonymised way. Each study participant had a username that was used throughout the study to collect the data and to refer to and identify the various study participants. The notifications were sent out every day at 5pm (see Supplementary Information S2.4 for further details on the notifications). Parallelism in the administration of the field-experiment with the two treatments, i.e. all subjects used the same app and were exposed to the same questionnaire etc., helped to maintain the excludability assumption, that is the potential outcome of the experiment depends solely on whether the subject receives the treatment [26].

3. Pilot Study Results

The collected data allows firstly to do some descriptive analyses, including looking at the trends in CO2 emissions throughout the study on the individual as well as aggregate level (see Supplementary Information S3.1). Generally, it is 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

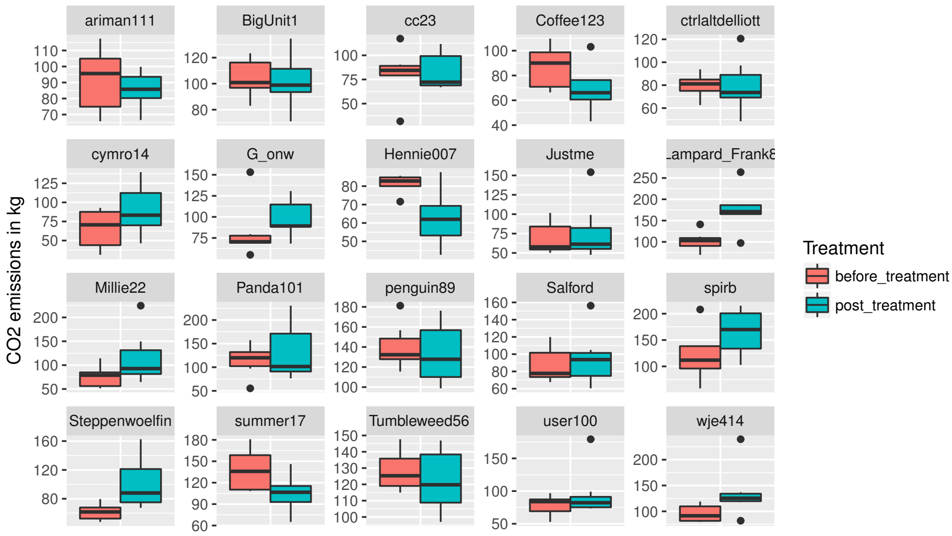

Fig. 1. Box plots for the 20 study participants showing their overall CO2 emissions before and after the field-experimental treatment.

today' (see Supplementary Information S3.1), this concern is not necessarily mirrored in their everyday environmental behaviour; 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 behaviour.

Of particular interest, at least in an actual study, is moreover to estimate the treatment effect and to compare various field-experimental interventions. Figure 1 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 behaviour [28], one week of intervention is not sufficient to measure consistent behavioural change. People cannot be expected to change their travelling 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.

However, if breaking down into the various environmental dimensions, we can see that it's not only the transport dimension that is inconsistent with our assumptions, in the food dimension for instance even a slightly significant negative treatment effect can be observed. The pilot study does not really allow 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 [29]. Only in the electricity dimension we can see a significant positive treatment effect (see Supplementary Information S3.2). 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' behaviour in positive (greater sustainability) way, except for the electricity dimension. However, as mentioned earlier, this was rather to be expected because behavioural adjustments takes usually more time [28]. 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 CO_2 emissions within each individual, which makes a paired t-test on individuals' average CO_2 emissions rather problematic.

This problem can be avoided to some extent in a repeated measure 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 (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 McFadden (0.02), Cox/Snell (0.17) and Nagelkerke (0.17) Pseudo-R-squares 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). Hence, we cannot necessary conclude that accounting for the treatments improves our ability to predict the overall CO_2 emissions. This might not be of great surprise given the limitation of the pilot study and the small data set. Nevertheless, when actually comparing the two treatments, the pilot study seems to suggests that the social monitoring treatment had a somewhat greater positive (in terms of reducing CO_2 emissions) effect on the overall environmental behaviour in comparison to the behavioural targeting treatment (see Figure 2). This result shows at least 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. Repeated measure ANOVA analyses for the different environment dimensions can be looked up in the Supplementary Information S3.3. They 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 bahavioural targeting treatment.

The data collected also allows to build and estimate choice models, either traditional discrete choice models [30] or if one would like to account for potential non-linearities in the decision making Gaussian Processes-based choice models suggested recently by Mann et al. [31, 32]. 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 to create a dataset of travel choices characteristics (e.g. CO_2 emissions) and account 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

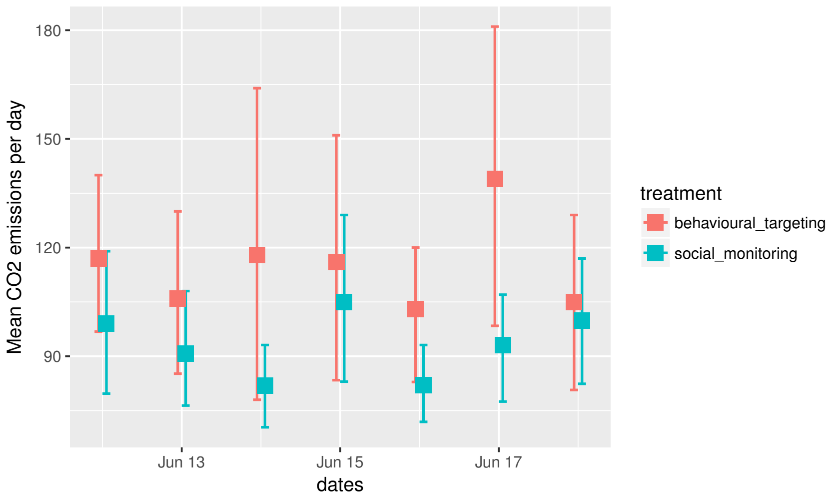

Fig. 2. Interaction plot shows the natural mean of each treatment*date combination along with the confidence interval of each mean with percentile method. Least Square Means for CO2 emissions in the behavioural targeting group were estimated to be 118.02 (se: 8.61, 95%-CI: [97.12, 138.93]), for social monitoring 93.19 (se: 8.52, 95%-CI: [72.42, 113.96]).

choices. Making use of Mann et al. [31] approach, allows to estimate utility functions to understand why study participants have chosen certain transport modes in a given situation. Keeping in mind the limitation of the data, i.e. we cannot draw any generalisable conclusions from these analyses, the obtained model results nevertheless show the potential of this approach to gain insights from an actual study.

Figure 3 shows for instance 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. This seems to interact to some extent with the financial situation. The transport mode choices of participants with more difficult financial situations 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 influenced by 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 models. 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 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) (see further results e.g. on the effect of transport cost

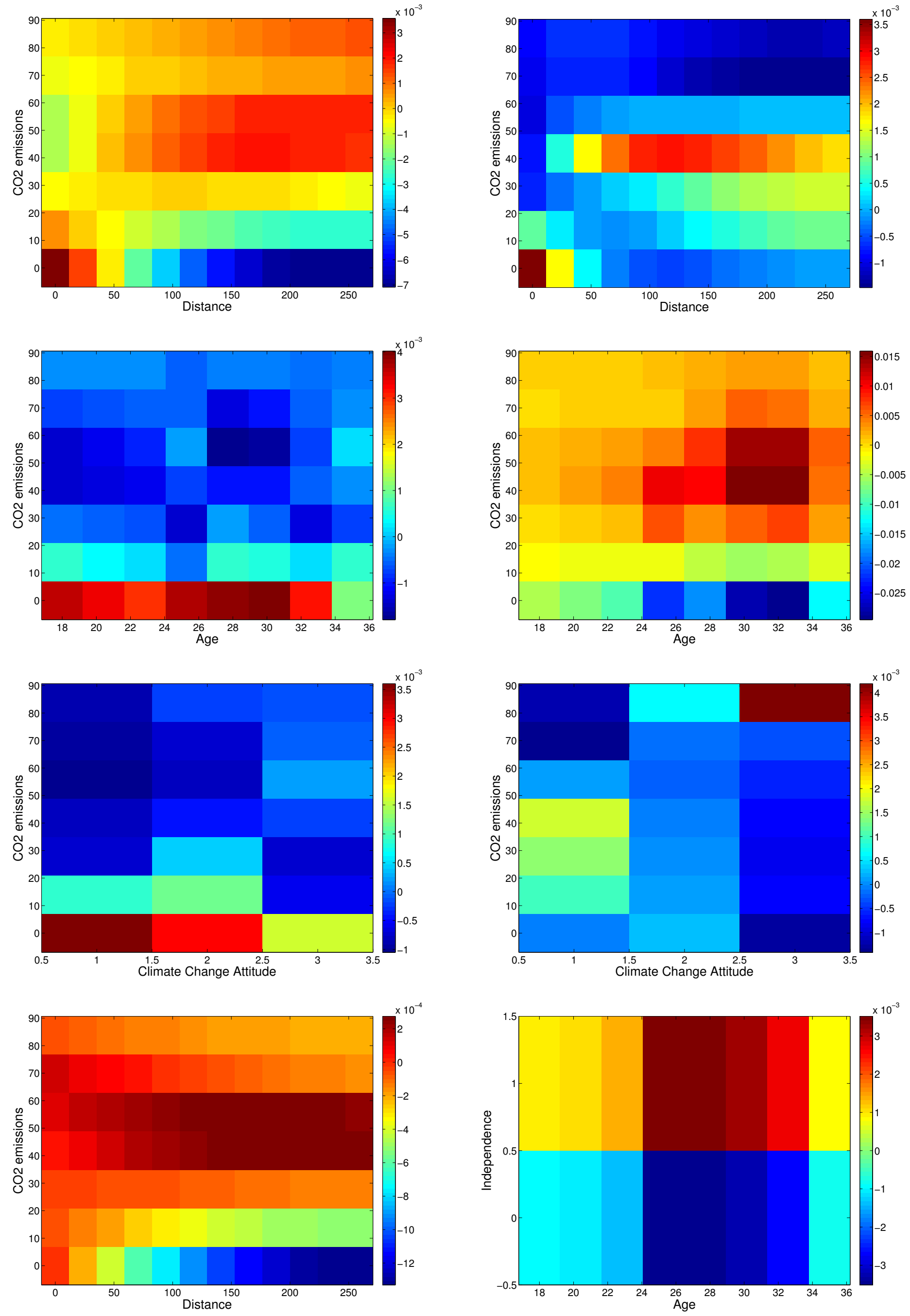

Fig. 3. Heat plots displaying the utility function for transport modes based on choice characteristics CO2 emissions and independence. The colour bar shows the utility scale, with redder colours indicating a positive utility and bluer colours none or even negative utility. The upper two panels show transport mode preferences with respect to travel distance for younger and financially well (left) or poor (right) participants (a similar pattern emerges for older participants, see Supplementary Information S3.4). The two panels in second row show the effect of age, holding financial situation constant at well, 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 participants who are financially well off.

in Supplementary Information S3.4). Though these results seem to be reasonable in describing people's behavioral choices, caution should be applied when interpreting these results, given the little data.

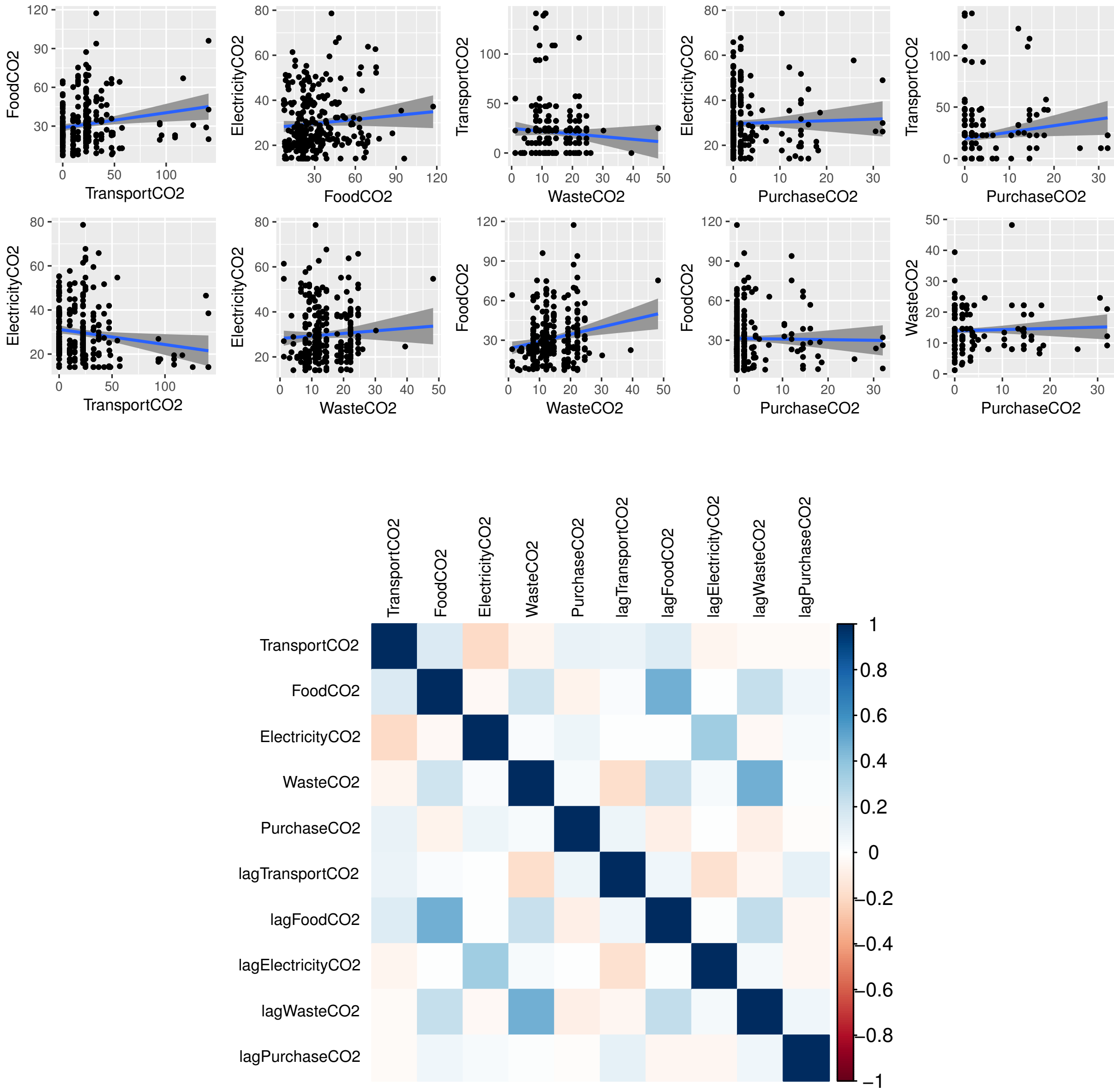

Fig. 4. The upper panels are scatterplots with a fitted linear regression line and 95%-Confidence Interval areas around it to display the bivariate relation between the various environmental dimensions. The bottom figure is a correlation matrix plot for second week data that displays the direction and strengths of correlations between the various environmental dimensions in colours, including one-day lags.

Finally, of interest is also to investigate how the various environmental behaviour dimensions interact, specifically whether we can find evidence for the so-called moral credential effect or self-licensing effect [33, 34], where a person who has chosen an environmentally friendly behaviour 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 behaviour 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 4). Weak evidence for compensation behaviour arises only in the relation between transport and electricity, where we see a negative correlation, though this might be an artefact, 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 4) 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 behaviour dimensions.

4. Lessons Learned from the Pilot Study - Future Research Outlook

Even though the results from the pilot study are in many ways non-conclusive due to various limitations of the pilot study, they nevertheless 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. The results show that even with the small pilot study sample we can observe some interesting behavioural patterns worth studying further. Moreover, 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 behavioural change, but there may be some potential for it.

It is not the intension of this study to suggest that the society should leave it to the individual responsibility of each citizen to fight climate change and strive for greater sustainability. Nudges etc. will not be sufficient to find a solution to the ecological crises humanity is facing. In fact one question that even this pilot study raises, is why are people not more environmentally friendly, even if they clearly think that sustainability and climate change is important? Gifford [35] suggests that “structural barriers such as a climate-averse infrastructure are part of the answer". Hence, policy measures such as CO2 emission taxes, public investment in sustainable infrastructure are inevitable if we seriously want to make a transition towards a sustainable future. But, the change on the individual level should be encouraged simultaneously with societal change. If both go hand in hand we are more likely to achieve a true transition. According to Gifford (ibid.), other factors that preclude people from acting are psychological barrier, such as "limited cognition about the problem, ideological world views that tend to preclude pro-environmental attitudes and behavior, comparisons with key other people, sunk costs and behavioral momentum, discredence toward experts and authorities, perceived risks of change, and positive but inadequate behavior change." We therefore need to study how change on the individual level can be achieved and the approach introduced here, could be one effective way to do that.

Of course this study approach too is not without limitations and problems. One limitation is the reliance on people's accurate reporting of their environmentally relevant activities. Here the approach suffers a weakness that most research involving humans is facing and there is no easy, obvious solution to this. At least the daily data collection makes sure that people don't have to struggle to remember what they did throughout the day. Moreover, the collection of more objective data, such as electric meter data etc. can to some extent allow to verify the responses. One way to try to achieve a higher accuracy is to gamify the data collection, so it's in people's interest to be accurate. The study design should be also improved in many other ways, some of which were mentioned already throughout the paper. Study participants' feedback in the final online survey provides valuable input for a better design among others of the survey questions. Besides developing a bespoke mobile phone application that solves the technical issues, it would be important to recruit a sizeable, preferably representative random sample for the study, to include a control group and to run the study over a much longer period of time, at least one month, to give study participants time to adjust their behaviour in response to interventions and to measure the consistency of behavioural change. Better data would allow to conduct better statistical analyses (e.g. mixed effect models, difference-in-difference analyses etc.). Furthermore, other interesting interventions could be included and the on tested here improved, e.g. in terms of behavioural targeting messaging [36] and in terms of unobtrusiveness [26]. Recent experimental studies in public good games show that reputation-based interventions are very effective in inspiring cooperative and prosocial behaviour [37, 38], therefore it would be worthwile including a reputation-based intervention in this study approach. This paper will hopefully inspire further thinking into how we can go a step further in studying social dilemmata and how we can overcome them.

Acknowledgements

Viktoria Spaiser would like thank the School of Politics and International Studies, University of Leeds, for SRIF (Strategic Research Investment Fund) funding that allowed to run this small pilot study.

References

[1] E. Oxtrom, T. Dietz, N. Dolsak, P. C. Stern, S. Stonich and E. U. Weber (eds.) *The Drama of the Commons*, National Academy Press, 2002.
[2] A. S. Pentland *Social Physics: How Good Ideas Spread - The Lessons from a New Science*, Penguin Press, 2014.
[3] H. Gijzen, Development: Big data for a sustainable future, *Nature*, **502** (2013). doi:10.1038/502038d.
[4] D. Giggs, M. Stafford-Smith, O. Gaffney, J. Rockström, M. C. Öhman, P. Shyamsundar, W. Steffen, G. Glaser, N. Kanie and I. Noble, Policy: Sustainable Development Goals for People and Planet, *Nature*, **495** (2013), 305–207. doi:10.1038/495305a.
[5] S. Hornibrook, C. May and A. Fearne, Policy: Sustainable Development and the Consumer: Exploring the Role of Carbon Labelling in Retail Supply Chains, *Business Strategy and the Environment*, **24** (2013), 266–276. doi:10.1002/bse.1823.
[6] C. Flath, D. Nicolay, T. Conte, C. van Dinther and L. Filopova-Neumann, Cluster Analysis of Smart Metering Data, *Business & Information Systems Engineering*, **4** (2012), 31–39. doi:10.1007/s11576-011-0309-8.
[7] M. Archetti and I. Scheuring, Coexistence of cooperation and defection in public goods games, *Evolution*, **65** (2010), 1140–1148. doi:10.1111/j.1558-5646.2010.01185.x.

[8] J. Jerit, J. Barabus and S. Clifford, Comparing Contemporaenous Laboratory and Field Experiments on Media Effects, *Public Opinion Quarterly*, **77** (2013), 256–282. doi:10.1093/poq/nft005.
[9] D. Rondeau and J. List, Matching and Challenges Gifts to Charity: Evidence from Laboratory and Natural Field Experiments, *Experimental Economics*, **11** (2008), 253–267. doi:10.3386/w13728.
[10] A. Biel, D. Elek, T. Gaerling and M. Gustaffson (eds.), *New issues and paradigms in research in social dilemmas*, Springer, 2008.
[11] A. D. I. Kramer, J. E. Guillory and J. T. Hancock, Experimental evidence of massive-scale emotional contagion through social networks, *PNAS*, **111** (2013), 8788–8790. doi:10.1073/pnas.1320040111.
[12] R. M. Bond, C. J. Fariss, J. J. Jones, A. D. I. Kramer, C. Marlow, J. E. Settle and J. H. Fowler, A 61-million-person experiment in social influence and political mobilization, *Nature*, **489** (2012), 295–298. doi:10.1038/nature11421.
[13] M. Tsvetkova, O. Nilsson, C. Öhmann, L. Sumpter and D. Sumpter, An Experimental Study of Segregation Mechanisms, *EPJ Data Science*, **5** (2016). doi:10.1140/epjds/s13688-016-0065-5.
[14] V. Pejovic and M. Musolesi, Anticipatory Mobile Computing for Behaviour Change Interventions, *Proceedings of the ACM International Join Conference on Pervasive and Ubiquitous Computing*, 2014, 295–298. doi:10.1145/2638728.2641284.
[15] Y. Zheng, S. V. Flanagan and A. Morabia, *Empowering Individuals to Make Environmentally Sustainable and Healthy Transportation Choices in Mega-Cities through a Smartphone App*, Final Report, University Transportation Research Center - Region 2, City University of New York, 2015, https://trid.trb.org/view.aspx?id=1371710
[16] S. Centellegher, M. De Nadai, M. Caraviello, C. Leonardi, M. Vescovi, Y. Ramadi, N. Oliver, F. Pianesi, A. Pentland, A. Fabrizio and B. Lepri,The Mobile Territorial Lab: a multilayered and dynamic view on parents' daily lives, *EPJ Data Science*, **5** (2016). doi:10.1140/epjds/s13688-016-0064-6.
[17] J. Chen and J. Stallaert, An Economic Analysis of Online Advertising Using Behavioral Targeting, *MIS Quarterly*, **38** (2014), 429–449.
[18] R. H. Thaler and C. R. Sunstain, *Nudge. Improving Decisions about Health, Wealth, and Happiness*, Penguin, 2008.
[19] S. Dawson, The impact of institutional surveillance technologies on student behaviour, *Surveillance & Society*, **4** (2006), 69–84.
[20] M. van Bommel, J.-W. van Prooijen, H. Elffers and P. A. M. van Lange, Be aware to care: Public self-awareness leads to a reversal of the bystander effect, *Journal of Experimental Social Psychology*, **48** (2012), 926–930. doi:10.1016/j.jesp.2012.02.011.
[21] B. Latané, The psychology of social impact, *American Psychologist*, **36** (1981), 343–356. doi:10.1037/0003-066X.36.4.343.
[22] 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.
[23] 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.
[24] 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.
[25] 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.
[26] A. S. Gerber and D. P. Green, *Field Experiments. Design, Analysis and Interpretation*, W.W. Norton & Company, 2012.
[27] M. Berners-Lee, *How Bad are Bananas? The Carbon Footprint of Everything*, Green Profile, 2010.
[28] 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.
[29] 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.
[30] M. E. Ben-Akiva and S. R. Lerman, *Discrete Choice Analysis: Theory and Application to Travel Demand*, MIT Press, 1985.
[31] 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*, Journal of Choice Modelling, forthcoming.
[32] R. P. Mann, *Gaussian process conditional logit model*, MATLAB code on GitHub, 2016, https://github.com/richardpmann/gp_conditional_logit.
[33] 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.
[34] 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.

[35] R. Gifford, The dragons of inaction: Psychological barriers that limit climate change mitigation and adaptation, *American Psychologist*, **66** (2011), 290–302. doi:10.1037/a0023566.
[36] R. B. Cialdini, Crafting Normative Messages to Protect the Environment, *Current Directions in Psychological Science*, **12** (2011), 105–109. doi:10.1111/1467-8721.01242.
[37] M. Milinski, D. Semmann and H.-J. Krambeck, Reputation helps solve the ‘tragedy of the commons’, *Nature*, **415** (2002), 424–426. doi:10.1038/415424a.
[38] F. Muñoz-Garcia, Competition for status acquisition in public good games, *Oxford Economic Papers*. doi:10.1093/oep/gpr007.