Data Science x (2018) 1–14
IOS Press

Advancing sustainability: using smartphones to study environmental behaviour 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

Abstract. Ecological sustainability is the defining challenge of our time. Here we suggest 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. Previous related studies are reviewed and we discuss how the boundaries of what can be done with smartphones for data collection and experimental purposes can be pushed further to allow for complex behavioural studies. We also present results from a pilot study to discuss the feasibility and potential of this approach. The pilot shows that studying social dilemma behaviour via smartphones is feasible and has potential value as an behavioural intervention tool.
Keywords: field-experiment, smartphone data, environmental behaviour, choice modelling, treatment effects

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 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.

*Corresponding author.

1. State of the Art

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 necessarily 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]. The largest bulk of studies using smartphones to collect data is to be found in health studies. In fact, a whole new area of research known as mobile health (mHealth) has emerged with the goal to identify behaviours that lead to positive or negative health outcomes in order to design and implement large-scale interventions [17]. Usually either smartphone usage data (e.g. call logs, short message service logs, app-use logs, battery-status logs, accelerometer, GPS, lights sensors, bluetooth scans, proximity sensors, voice, etc.) is collected that provides information about peoples behavioural lifestyles [18, 19] or bespoke (self-monitoring) software applications are developed that allow researchers to collect specific data. [20] for instance developed and used a self-monitoring smartphone software "MONARCA" to collect data on physical and social activities over three months from patients with bipolar disorder to predict depression episodes. While [21] used a smartphone application with certain tasks to be performed by participants to detect signs of parkinson by assessing voice, posture, gait, finger tapping and response time. Studies with bespoke (self-monitoring) software applications are however rather rare, most studies collect sensory or smartphone usage data.

Most of these studies however do not implement an experimental design. In particular smartphone-sensing studies, where smartphone usage data is collected passively and automatically in the background produce purely observational data, which may provide interesting insights e.g. on people's lifestyles, well-being, performance, social interactions [22] etc. but which is insufficient if one is interested in causal mechanisms. The raw collected sensory data requires moreover heavy and sophisticated processing to infer behaviours of interest [23, 24] and it is often far from clear how reliable and valid these behavioural inferences are [19]. Other studies combine sensing with self-reported data e.g. of subjective experience, often obtained through ecological-momentary assessments (EMAs), which can be context-contingent (e.g. GPS record triggers question: "What are you doing here?" [25, 26]). But even these studies do not use an experimental setup, although experimental sensing-data collection apps have been developed such as RecordMe [27] or UBhave [28]. On the other hand, some studies try to make causal inferences from quasi-experimental frameworks [29]. Studies that involve experiments are usually studies where the smartphone application itself or some other system is tested for its efficacy [30], rather than where the smartphone application is used to issue an experimental intervention in the data collection process. Nevertheless, the utilisation of mobile technologies for experimental treatments is increasingly discussed and tentatively explored, particularly in medical research [31]. All studies ultilized SMS and/or other multimedia message services for interventions and about half of them reported significant health behavioural changes (ibid.).

The increasing utilisation of smartphone technologies for data collection and intervention goes along with another trend. The field experimental approach is a rapidly growing form of social science research, encompassing hundreds of studies on topics like education, crime, employment, poverty, development, discrimination, political participation etc. [32, 33]. And the smartphone-based experimental intervention approach can be well implemented within a field experiment study, indeed turning these mobile devices into real-life laboratories.

While using smartphones to collect data is now becoming quite common in health and psychological studies, there are no studies where smartphones are used to study complex social and/or choice behaviours, such as behaviours in social dilemma or public good situations. There is no straightforward way to infer such complex behaviours from sensing or smartphone usage data, given it is even non-trivial to infer much simpler behavioural patterns (e.g sleeping vs. being awake) from such data [19]. Ecological-momentary assessments (EMAs) would have to be ulitised within bespoke software applications. And if a field-experimental approach is chosen the question is how to best translate the various experimental interventions tested in laboratory experiments on public good and social dilemmas into smartphone-based interventions. There is one study from the US that at least to some extent leads the way. [34] developed a bespoke smartphone application onTrac and assessed the impact of behavioural nudge interventions implemented in the app in a randomized controlled trial. Specifically, onTrac reported carbon emissions and calories burned associated with user specified travel modes as an intervention. An in-build accelerometer detected automatically some of the travel models (walking and bicycling) automatically using GPS records to estimate users' speed, while the users had to report other types of travel modes (e.g. car, bus, train, subway). User survey following a three week trial of onTrac app usage revealed increases in self-reported considerations for the environmental impact of travel choices among students who used onTrac comparing to a control group who did not. While this intervention is certainly interesting – though it's problematic to mix a environmental awareness and a health awareness nudge, because it is impossible to know which of them has what effect –, previous laboratory experimental research has produced a whole set of other interesting interventions in the public goods context that would be worth trying in a real-life setting. Moreover, while a focus on transport behaviour is certainly sensible in the context of the study, environmental behaviour is necessarily multidimensional, involving transport, energy consumption, food choices, waste behaviour etc. And these various behavioural dimensions do most likely interact in real life.

2. Using smartphones to study (environmental) social dilemma problems in a field-experimental setup

This paper suggests to study environmental behaviour (transport behaviour, energy consumption, food consumption, goods consumption, wasting behaviour) in real life by using smartphones to collect daily behavioural data over an extended period of time in a field-experimental setup. The major originality of the approach suggested here and partly tested in a small 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. Overcoming environmental behaviour dilemmas is essential to a successful transition to an ecologically sustainable society, as envisioned by the United Nations. While pollution and depletion of natural resources is a prime example for social dilemmas, they can be encountered in all areas of life where collective action is required: civil society relies on volunteering, democracy relies on active democratic participation, public spaces rely on peoples' other-regarding behaviour etc. Given the pervasiveness of social dilemmas, it is not surprising that social dilemmas are one of the core human behaviour research problems [1] and a core issue in every society. Any proven success in understanding how social dilemma problems can be dealt with in real life situations could therefore have far reaching consequences and allow a translation of the research results into policy measures. Today's data-generating digital technologies offer new possibilities to study human behaviour in real life social dilemma situations.This has been recognised by the UN, which established an Independent Expert Advisory Data Revolution Group to make concrete recommendations on bringing about a data revolution in sustainable development (http://www.undatarevolution.org).

One boundary of what can be done with "living laboratory" data that we suggest to push is the complexity and multi-dimensionality of the behaviour measured. We suggest to study multiple environmental behaviour dimensions (transport behaviour, energy consumption, food consumption, goods consumption, wasting behaviour) simultaneously in order 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 [35, 36], 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.

Collecting data on multiple environmental behaviour and issuing experimental interventions requires a bespoke smartphone application software solution. Sensing or smartphone usage data will not be sufficient, though could be partly used to complement ecological-momentary assessments (EMAs), essentially questions that require users to self-report behaviour (e.g. what food was consumed, what waste was produced with some default answers to make the data entry quicker). Thus, data to be collected would include text input. Typing answers could however be potentially replaced through voice recording of answers. These voice recordings could then be automatically converted to text data. Furthermore, barcode scans should be collected to assess goods consumption. These barcode scans data needs to be linked to a barcode database, ideally to one that contains information about how sustainable the respective product is. Potentially, the barcode scanning function could be linked to smart barcode scanner applications like GoodGuide that allow the automatic identification of environmentally friendly products. Furthermore data on electricity usage should be collected. Users could take a picture of the electric meter counter with their smartphones and OCR (Optical Character Recognition) algorithms could be used to extract the number from the picture. This solution could be implemented for users who do not have a smart meter. Users with smart meters could access the data from their smart meters through their smartphones and give permission to the data collection application to access that data. However, at this stage smart meters are still not very common. Image taking could potentially also replace text input on foods consumed. AI-based smartphone applications are being developed now, which can translate food images into a list of ingredients [37]. Users could then just correct potential errors (e.g. soya burger instead of a meat burger). Furthermore, GPS records should be obtained to estimate travel distances and potentially to infer transportation modes [38]. It is, thus, suggested to combine sensing data with ecological-momentary assessments data, whereby the sensing data could be partly used to verify self-reported behaviour. The challenge is to find a way to collect quickly and effectively sufficient and insightful data, without interfering too much with users' everyday life. Generally, users should be able to record data anytime throughout their day. The captured data could then be translated for instance into average CO2 emissions, based on calculations provided for instance by [39]. Thus, though various behaviour is recorded, all these behaviours are quantified in terms of environmental impact by a common measure. Further data could be potentially collected on reasons for respective behaviours, e.g. why a certain transport mode was chosen, or who else was involved.

Another boundary to be pushed is on the experimental interventions. Presently, studies typically implement only a single experimental intervention and almost exclusively it is message-based intervention, where study participants are nudged to display a particular behaviour. Other potentially interesting interventions have not been studied yet in a field-experiment setup using smartphones for data collection and experimental intervention. We suggest to study multiple experimental interventions that could provide information on which intervention are most effective in terms of real-life behavioural change. Laboratory studies on public goods dilemmas can be very instructive in designing such interventions. Message-based interventions could be further developed through behavioural targeting, originally an online advertisement practice where online users are presented with advertisement based on their past online behaviour [40]. Behavioural targeting is based in the nudge theory [41]. 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.). I.e. Nudges could be tailored messages sent to study participants' smartphones, proposing specific behavioural changes based on participants' past behaviour.

Another interesting intervention to be studied is social monitoring. Individuals in this treatment group 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 [42, 43]. This hypothesis is based in the social influence theory which investigates the effects of compliance, conformity and competition [44].

Other field-experimental treatments, e.g. reputation-based interventions or financial incentives could be implemented in an actual study too. Particularly the reputation mechanism proved quite effective in solving the tragedy of the commons, at least in laboratory public good games [45]. A reputation-based intervention could be designed for instance by dividing the study participants into two competing groups, whereby both groups have to goal to collectively reduce their CO2 emissions. In each group the study participants would be given the opportunity to monitor each other (as in the social monitoring group) and to communicate with each other through messaging implemented in the smartphone application. Individuals will be ranked in each group based on their contribution to reducing CO2 emissions and this rank would be visible to everyone in the group through the smartphone application. [46] found for instance in a large-scale field experiment that sufficiently high observability promotes cooperation in public good games much more effectively than financial incentives.

One could even go a step further, though that would require greater and/or additional research into the software engineering and artificial intelligence (AI) side. Many of the interventions described here are rather patronizing and to some extent may be seen as manipulative. But how could we encourage behavioural change in a more emacipatory, empowering way? Should the goal not be to empower individuals in their capability to make informed decision that are right for them and the society overall? But how could an emancipatory intervention look like? Our idea here is to design an AI-based decision-making assistant that would help users to make decisions that are right for them and the society overall.

This AI-based assistant could provide the users with all necessary information (e.g. what the options are, what the costs are financially and in terms of CO2 emissions) and answer their questions (e.g. what are the alternatives), additionally it could engage the users in a Socratic dialogue [47] encouraging them to think critically and make an autonomous and yet responsible decision, considering society's greater good and question their preferences and habits. This intervention could be tested against a "non-emancipatory" AI-based decision-making assistant which would also provide the user with all necessary information and answer their questions, but which would then suggest to participants an "optimal" decision solution based on participants' stated preferences.

Generally, the experimental design of the research project will make it necessary to analyse the data partly on the run (quasi real-time) during the data collection phase. That is, data analysis at this stage of the project will be part of the data collection and field experiment process, to compute for instance environmental behaviour scores and to visualise collected data in order to issue the respective experimental interventions.

Finally, the suggested approach should certainly follow a Randomised Controlled Trial (RCT) field-experimental design to allow for causal inference in hypothesis testing. The data collection and field-experimental interventions should take place over a longer period of time, at least one month, to give study participants time to respond to experimental interventions and to test, whether observed behavioural changes are stable over time.

3. Pilot Study

To show the feasibility of the above described approach in general, a pilot study was conducted over two weeks in June 2017 with 20 study participants. Two field-based interventions were tested to inspire co-operative, i.e., environmentally-friendly behaviour: (1) behavioural targeting and (2) social monitoring. A control group was not included in the pilot study due to financial restriction that allowed to recruit and compensate only 20 study participants. The primary goal of the pilot was to show the feasibility of the study approach (incl. multiple interventions), consequently, it was decided to implement a second intervention rather than a control group. Instead, the field-experimental treatments were only issued in the second week of the study. The first week thus served as a reference point for comparison and treatment effect estimation. In an actual study however a control group is absolutely indispensable, also a sizeable, representative sample should be aimed for in an actual study.

3.1. Data Collection

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 [48, 49]. The platform allows to create project-specific smartphone applications and then publish these through the EpiCollect 5 mobile phone application, that operates on iOS and Android smartphones. 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 gives users full control over their data; they have to explicitly upload the data.

EpiCollect 5 is a great platform for research data collections but it is not designed for experimental research. For an actual study on larger scale a bespoke software solution would be preferable, which would facilitate certain features more directly (see Supplementary Information S1 for further discussion on smartphone application, including ongoing work on a bespoke solution).

From the 20 recruited study participants, 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 [32] e.g. through a spill-over effect between study participants who are friends. This was done by recruiting students from different disciplines, graduation levels, courses etc.

Given data collection tool place in June, students were on term break 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. The recorded data (e.g. answers on what transport mode was used, what food consumed, what electronic devices used, etc.) was translated into average CO2 emissions for the specified activity based on [39]. This allowed to calculate average CO2 emissions for each environmental behaviour dimension and overall (see Supplementary Information S2.1 for further discussion of data collection, including questionnaire implemented in the app).

At the start of the pilot study the study participants were 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). At 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, whether they thought that participation in the study raised their environmental awareness) (see Supplementary Information S2.2. and S2.3 for further details).

After the first week the 20 study participants were randomly assigned to one of the two field-experimental groups, each containing 10 study participants. In the second week all study participants were subject to one of the two 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 receive 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 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 [32].

3.2. Results

The results of pilot data analysis will be discussed 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. 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.

The descriptive analyses of the collected data, including looking at the trends in CO2 emissions throughout the study on the individual as well as aggregate level can be accessed in the Supplementary Information S3.1. Moreover, for a full discussion of treatment effects (including the rather inconclusive results from t-tests) please see Supplementary Information 3.2. Here only some selected results will be presented. Among others we run a repeated measure ANOVA, that allowed us to compare the two field-experimental interventions, 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). 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 suggests that the social monitoring treatment had a somewhat greater positive (in terms of reducing CO2 emissions) effect on the overall environmental behaviour in comparison to the behavioural targeting treatment (see Figure 1). This result shows at least that it could be worth investigating further the effects of different types of treatments in a full study, including a control-group and potentially other treatments.

The collected data also allows for building and estimating choice models, either traditional discrete choice models [50] 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. [51, 52] are quite useful. 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 was linked to the travel characteristic choices. Making use of the Mann et al. [51] 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 2 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. This seems to interact to some extent with the financial situation. The transport mode choices of less affluent

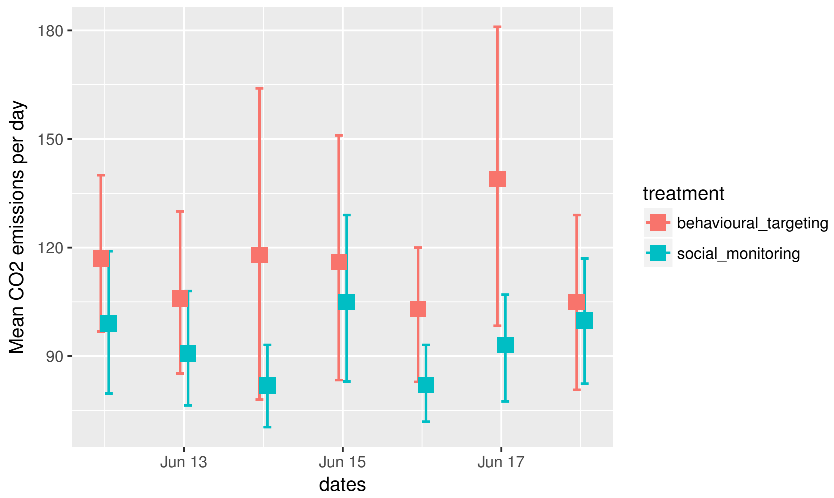

Fig. 1. Interaction plot shows the natural mean of each treatment*date combination along with the confidence interval of each mean with percentile method. Least Square Means for CO2 emissions in the behavioural targeting group were estimated to be 118.02 (se: 8.61, 95%-CI: [97.12, 138.93]), for social monitoring 93.19 (se: 8.52, 95%-CI: [72.42, 113.96]).

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 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 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.

Analyses have also been conducted 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 discussed earlier. Please see Supplementary Information S3.5 for a full account of the results. In summary, the analyses did not result in clear interaction patterns between the various environmental dimensions. A weak relation could only be established between the transport and electricity dimension. The small data set and other pilot study limitations may have inhibited measuring clearer relations.

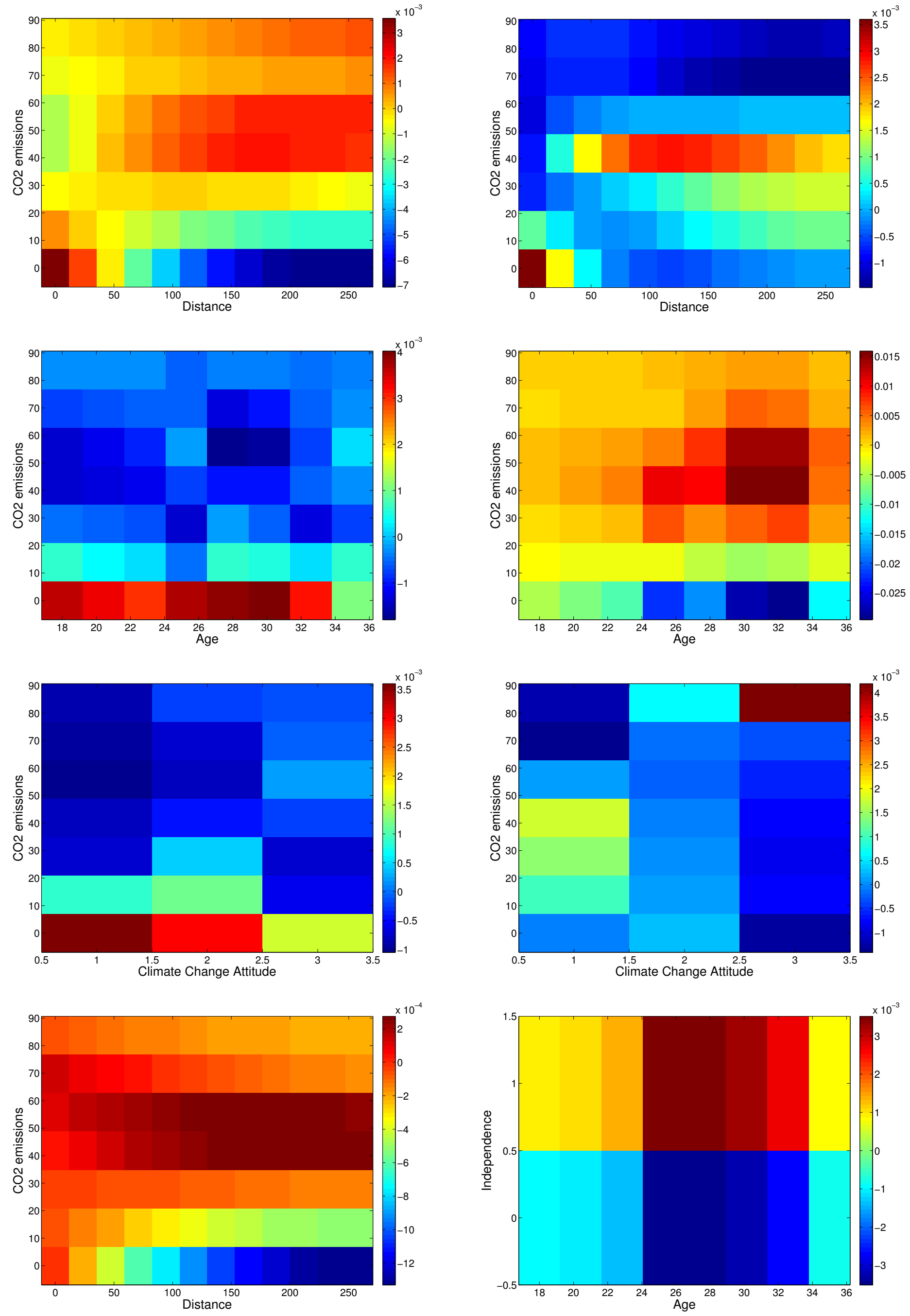

Fig. 2. Heat plots displaying the utility function for transport modes based on choice characteristics CO2 emissions and independence. The colour bar shows the utility scale, with redder colours indicating a positive utility and bluer colours none or even negative utility. The upper two panels show transport mode preferences with respect to travel distance for younger and financially well (left) or less affulent (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.

3.3. Discussion

The pilot study results presented in the main manuscript 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. Furuthermore, 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.

The reliance on people's accurate reporting of their environmentally relevant activities is problematic. 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 for response verification. When designing follow-up studies more thought should go into further automatising data collection, e.g. usage of accelerators for transport mode inference [40].

There is also much room for other improvements of the study design tested in the pilot, as already mentioned earlier (e.g. in terms of sampling, duration of the study, control group etc.). Study participants' feedback in the final online survey provides valuable input for a better design among others of the survey questions. For instance, sometimes the type and usage duration of certain devices like lamps, or the type and amount of certain foods (e.g. organic, local vegetables) can be more indicative of the environmental implications of the behaviour. The problem of including more details however is of course that the time participants spend on providing the answers would increase. Furthermore, besides including other interesting interventions, the ones tested in the pilot could be improved too, e.g. in terms of behavioural targeting messaging [53] and in terms of unobtrusiveness [32]. Nevertheless, the pilot study shows that studying (environmental) social dilemmas via a smartphone in a field-experimental setup is feasible and could lead to new insights.

Finally, it should be noted that it is not the intention 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. [54] suggests for instance that "structural barriers such as a climate-averse infrastructure are part of the answer" why people who are environmentally concerned do not necessary act more environmentally friendly. Hence, policy measures such as taxing companies for CO2 emission, public investment in sustainable infrastructure etc. 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.

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. Ostrom, T. Dietz, N. Dolsak, P.C. Stern, S. Stonich and E.U. Weber (eds.), *The Drama of the Commons*, National Academy Press, 2002. doi:10.17226/10287.

[2] A.S. Pentland, *Social Physics: How Good Ideas Spread - The Lessons from a New Science*, Penguin Press, 2014. ISBN: 1594205655

[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. doi:10.1007/978-0-387-72596-3.

[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 Joint Conference on Pervasive and Ubiquitous Computing, Seattle, USA*, (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] N. Lathia, V. Pejovic, K.K. Rachuri, C. Mascolo, M. Musolesi and P.J. Rentfrow, Smartphones for large-scale behaviour change interventions, *IEEE Pervasive Computing*, **12** (2013), 66–73. doi:10.1109/MPRV.2013.56.

[18] G. Chittaranjan, J. Blom and D. Gatica-Perez, Mining large-scale smartphone data for personality studies, *Personal and Uniqitous Computing*, **17** (2013), 433–450. doi:10.1007/s00779-011-0490-1.

[19] G.M. Harari, N.D. Lane, R. Wang, B.S. Crosier, A.T. Campbell and S.D. Gosling, Using Smartphones to Collect Behavioural Data in Psychological Science: Opportunities, Practical Considerations, and Challenges, *Perspect Psychological Science*, **11** (2017), 838–854. doi:10.1177/1745691616650285.

[20] M. Faurholt-Jepsen, M. Frost, M. Vinberg, E.M. Christensen, J.E. Bardram and L.V. Kessing, Smartphone data as objective measures of bipolar disorder symptoms, *Psychiatry Research*, **217** (2013), 124–127. doi:10.1016/j.psychres.2014.03.009.

[21] A. Arora, V. Venkataraman, A. Zhan, S. Donohue, K.M. Biglan, E.R. Dorsey and M.A. Little, Detecting and monitoring the symptoms of Parkinson's disease using smartphones: A pilot study, *Parkinson & Related Disorders*, **21** (2015), 650–653. doi:10.1016/j.parkreldis.2015.02.026.

[22] R. Wang, F. Chen, Z. Chen, T. Li, G. Harari, S. Tignor and A.T. Campbell, Studentlife: Assessing mental health, academic performance and behavioural trends of college students using smartphones, *Proceedings of the 2014 ACM International Joint Conference on Pervasive and Ubiquitous Computing, New York, USA*, (2014), 3–14. doi:10.1145/2632048.2632054.

[23] Z. Chen, M. Lin, F. Chen, N.D. Lane, G. Cardone, R. Wang, ... and A.T. Campbell, Unobstrusive sleep monitoring using smartphones, *Proceedings of the 7th International Conference on Pervasive Computing Technologies for Healthcare: PervasiceHealth, Brussels, Belgium*, (2013), 145–152. doi:10.4108/icst.pervasivehealth.2013.252148.

[24] S. Saeb, M. Zhang, C.J. Karr, S.M. Schueller, M.E. Corden, K.P. Kording and D.C. Mohr, Mobile phone sensor correlates of depressive symptom severity in daily-life behaviour: An exploratory study, *Journal of Medical Internet Research*, **17** (2015), e175. doi:10.2196/jmir.4273.

[25] K.K. Rachuri, M. Musolesi, C. Mascolo, P.J. Rentfrow, C. Longworth and A. Aucinas, EmotionSense: A mobile phones based adaptive platform for experimental social psychology research, *Proceedings of the 12th ACM International Conference on Ubiquitous Computing, New York, USA*, (2010), 281–290. doi:10.1145/1864349.1864393.

[26] V. Pejovic, N. Lathia, C. Mascolo and M. Musolesi, Mobile-based experience sampling for behaviour research, In: M. Tkalčič, B. De Carolis, M. de Gemmis, A. Odić, A. Košir, *Emotions and Personality in Personalized Services*, Springer, (2016), 141–161. doi:10.1007/978-3-319-31413-6_8.

[27] D. Blachon, F. Portet, L. Besacier and S. Tassart, RecordMe: A Smartphone Application for Experimental Collection of Large Amount of Data Respecting Volunteer's Privacy, *Proceedings of the 8th International Conference on Ubiquitous Computing and Ambient Intelligence, Belfast, UK*, (2014), 345–348. doi:10.1007/978-3-319-13102-3_57.

[28] C. Hargood, V. Pejovic, L. Morrison and L. Yardley, The UBhave Framework: Developing Dynamic Mobile Applications for Digital Behavioural Interventions, 2016. https://eprints.soton.ac.uk/372016/1/poster-ab.pdf.

[29] F. Tsapeli and M. Musolesi, Investigating causality in human behaviour from smartphone sensor data: a quais-experimental approach, *EPJ Data Science*, **4** (2015), 24. doi:10.1140/epjds/s13688-015-0061-1.

[30] P.M. Enock, S.G. Hoffmann and R.J. McNally, Attention Bias Modification Training Via Smartphones to Reduce Social Anxiety: A Randomized Controlled Multi-Session Experiment, *Cognitive Therapy and Research*, **38** (2017), 200–216. doi:10.1007/s10608-014-9606-z.

[31] E.R. Buhi, T.E. Trudnak, M.P. Martinasek, A.B. Oberne, H.J. Fuhrmann and R.J. McDermott, Mobile phone-based behavioural interventions for health: A systematic review, *Health Education Journal*, **72** (2013), 564–583. doi:10.1177/0017896912452071.

[32] A.S. Gerber and D.P. Green, *Field Experiments. Design, Analysis and Interpretation*, W.W. Norton & Company, 2012. ISBN: 9780393979954.

[33] D.P. Green and A.S. Gerber, *Get out the Vote! How to Increase Voter Turnout*. 2nd ed., Brookings Institution Press, 2008. ISBN: 9780815725688.

[34] 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, City University of New York, 2015. https://www.utrc2.org/sites/default/files/pubs/Smart-Transportation-Choices-Mega-Cities-Smartphone-App.pdf.

[35] 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.

[36] 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.

[37] A. Conner-Simons and R. Gordon, *Artificial intelligence suggests recipes based on on food photos*, MIT News, 2017. http://news.mit.edu/2017/artificial-intelligence-suggests-recipes-based-on-food-photos-0720.

[38] S. Dabiri and K. Heaslip, Inferring transportation modes from GPS trajectories using a convolutional neural network, *Transportaton Research Part C*, **86** (2018), 360–371. doi:10.1016/j.trc.2017.11.021..

[39] M. Berners-Lee, *How Bad are Bananas? The Carbon Footprint of Everything*, Green Profile, 2010. ISBN: 1846688914.

[40] J. Chen and J. Stallaert, An Economic Analysis of Online Advertising Using Behavioral Targeting, *MIS Quarterly*, **38** (2014), 429–449. doi:10.25300/MISQ/2014/38.2.05..

[41] R.H. Thaler and C.R. Sunstain, *Nudge. Improving Decisions about Health, Wealth, and Happiness*, Penguin Press, 2008. ISBN: 0141040017

[42] S. Dawson, The impact of institutional surveillance technologies on student behaviour, *Surveillance & Society*, **4** (2006), 69–84. https://ojs.library.queensu.ca/index.php/surveillance-and-society/article/view/3455.

[43] 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..

[44] B. Latané, The psychology of social impact, *American Psychologist*, **36** (1981), 343–356. doi:10.1037/0003-066X.36.4.343.

[45] 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..

[46] E. Yoeli, M. Hoffman, D.G. Rand and M.A. Nowak, Powering up with indirect reciprocity in a large-scale field experiment, *PNAS*, **110** (2013), 10424–10429. doi:10.1073/pnas.1301210110..

[47] Y. Goda, M. Yamada, H. Matsukawa, K. Hata and S. Yasunami, Conversation with a Chatbot before Online EFL Group Discussion and the Effects of Critical Thinking, *The Journal of Information and Systems in Education*, **13** (2014), 1–7. doi:10.12937/ejsise.13.1..

[48] 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.

[49] 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.

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

[51] 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*, PLoS ONE, forthcoming, preprint on SSRN: https://papers.ssrn.com/sol3/papers.cfm?abstract_id=2869215

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

[53] 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.

[54] 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.