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

Viktoria Spaiser [a,*], Davide Luzzatti [b] and Aliki Gregoriou [c]

[a] *School of Politics and International Studies, University of Leeds, LS2 9JT, Leeds, The United Kingdom*
E-mail: v.spaiser@leeds.ac.uk; ORCID: https://orcid.org/0000-0002-5892-245X
[b] *Leeds University Business School, University of Leeds, LS2 9JT, Leeds, The United Kingdom*
E-mail: dadelutz@gmail.com; ORCID: https://orcid.org/0000-0003-2985-8302
[c] *School of Computing, University of Leeds, LS2 9JT, Leeds, The United Kingdom*
E-mail: alikigreg@gmail.com; ORCID: https://orcid.org/0000-0002-5910-4889

Abstract. Ecological sustainability is the defining challenge of our time. Here we suggest a methodological approach that could help to investigate how environmental behavior (transport behavior, energy consumption, food consumption, goods consumption, wasting) dilemmas can be overcome on an individual level in real life by using smartphones to collect daily behavioral 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 behavioral studies. Results from a pilot study are presented to discuss the feasibility and potential of this approach. The pilot shows that studying social dilemma behavior via smartphones is feasible and has potential value as a behavioral intervention tool.

Keywords: Field-experiment, smartphone data, environmental behavior, social dilemma, sustainability, living laboratory

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 face is how to make the transition towards a sustainable

---

*Corresponding author. E-mail: v.spaiser@leeds.ac.uk.

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 behavior [40]. This paper suggests to study environmental behavior in real life social dilemma situations by exploiting smartphone technology to collect new types of "living laboratory" [43] 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 behavior in real life.

2. State of the art

Big data is widely regarded as a rich data source for (environmental) human behavior [23,24]. Typically, consumer behavior is the focus of environmental behavior studies making use of big data such as retailers' loyalty cards data [30] or smart meter data [20]. However, such data is limited. For instance, Hornibrook et al. [30] 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. They 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 [3]. But, studies have also shown that the correspondence between laboratory experimental and field-experimental results is often quite weak [31,46], suggesting that we cannot necessarily make conclusions about real life (social dilemma) behavior from laboratory experiments. Consequently, there is a lack of deeper theoretical understanding of how these dilemma mechanisms play out in real life [6] beyond non-generalizable case studies [40].

The most recent methodological developments aim to combine the big data approach with an experimental design [8,35,50]. Mobile technologies can be ideal tools for such combined approaches [42,50,55].

2.1. Data collection via smartphones

Smartphones are increasingly used to study people's daily lives, tracking among others social interactions or mobility routines [10]. 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 behaviors that lead to positive or negative health outcomes in order to design and implement large-scale interventions [37]. Usually 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 people's behavioral lifestyles [13,28]. On the other hand, bespoke (self-monitoring) software applications are developed that allow researchers to collect specific data. Smartphone-sensing studies, where smartphone usage data is collected passively and automatically in the background, produces purely observational data, which may provide interesting insights e.g. on people's lifestyles, well-being, performance or social interactions [52]. But, such data is problematic if one is interested in causal mechanisms. Though, there are studies that make causal inferences from quasi-experimental frameworks [49].

The raw collected sensory data requires moreover heavy and sophisticated processing to infer behaviors of interest [12,47] and it is often far from clear how reliable and valid these behavioral inferences are [28]. Other studies combine sensing with self-reported data e.g. on 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?" [27,41,45]). And there are studies that rely entirely on smartphone-based self-monitoring. [19] 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 a bipolar disorder to predict depression episodes. And [4] used a smartphone application featuring certain tasks to be completed by participants in order 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.

2.2. Experimental designs in smartphone-based studies

Some smartphone-based studies implement an experimental design. Experimental sensing-data collection apps have been specifically developed for such purpose, such as RecordMe [7] or UBhave [29]. Studies that involve experiments are often studies where the smartphone application/some IT system itself [18] is tested or where the "intervention" is in the external environment [27]. Less common are studies where the smartphone application is used to issue an experimental intervention in the data collection process in order to incite behavioral changes. However, the utilization of mobile technologies for experimental treatments is increasingly discussed and explored, particularly in medical research [9,33]. Most studies utilize SMS and/or other multimedia message services for interventions and about half of them reported significant health behavioral changes [9]. But, there are a few studies that implement a more complex and sophisticated intervention design. An example is the Active2Gether app that not only intervenes through coaching messages, but, additionally facilitates social comparison to encourage physical activity. Based on Facebook friendship networks the app provides group averages on the group level and ranks user's performance within a list of other users on the individual level [33].

The increasing utilization 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 or political participation [21,26]. 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.

2.3. Smartphone-based studies of environmental behaviour

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 behaviors, such as behaviors in social dilemma or public good situations. There is no straightforward way to infer such complex behaviors from sensing or smartphone usage data, given it is even non-trivial to infer much simpler behavioral patterns (e.g. sleeping vs. being awake) from such data [28]. Ecological-momentary assessments (EMAs) would have to be utilized 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 to some extent leads the way. [56] developed a bespoke smartphone application onTrac and assessed the impact of behavioral 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 modes (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). A 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, 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, the study is problematic, because it mixes an environmental awareness and a health awareness nudge and this makes it impossible to know which of them had what effect. Finally, while a focus on transport behavior is certainly sensible in the context of the study, environmental behavior is necessarily multi-dimensional, involving for instance transport, energy consumption, food choices, waste behavior. And these various behavioral dimensions do most likely interact in real life.

Further inspiration can be drawn from other technology-based intervention platforms and studies, of which some are still work in progress. [34] for instance suggest the ASSET pilot platform for sustainable (food) consumerism, consisting of an ASSET app, ASSET supermarket-based localization system and an ASSET database system. The idea behind ASSET is to allow users to follow their personal sustainability preferences when making purchases by providing them with objective, yet tailored product information. Though ASSET is certainly a very interesting and promising platform, one may question whether personal sustainability preferences are as stable and reliable as the engineers of this system seem to assume. And is it the lack of (objective) information that prevents people from acting more sustainably and hence is an information-based intervention the most effective intervention? If behavioral change is the goal then various interventions should be tested against each other.

3. Using smartphones to study environmental social dilemma problems in a field-experimental setup

The approach suggested here and partly tested in a small pilot project intends to push the boundaries of what can be done with "living laboratory" data in order to better understand (environmental) social dilemma problems. Specifically, this paper suggests to study environmental behavior (transport behavior, energy consumption, food consumption, goods consumption, wasting behavior) in real life by using smartphones to collect daily behavioral data over an extended period of time in a field-experimental setup. A Randomized Controlled Trial (RCT) field-experimental design should be adopted to allow for causal inference in hypothesis testing. Moreover, the data collection and field-experimental interventions should take place over a long period of time, at least one month, to give study participants time to respond to experimental interventions and to test, whether observed behavioral changes are stable over time.

Overcoming environmental behavior dilemmas is essential for 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 behavior etc. Given the pervasiveness of social

dilemmas, it is not surprising that social dilemmas are one of the core human behavior research problems [40] 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 behavior in real life social dilemma situations.This has been recognized 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).

3.1. Collecting multi-dimensional environmental behavior data

One boundary of what can be done with "living laboratory" data (i.e. big data collected in a field-experimental setting) that we suggest to push is the complexity and multi-dimensionality of the behavior measured. We suggest to study multiple environmental behavior dimensions (transport behavior, energy consumption, food consumption, goods consumption, wasting behavior) simultaneously in order to understand how they interact in people's decision making, e.g. when people decide to buy an ecological product to compensate for environmentally damaging travel behavior. Including different environmental behavior dimensions allows to investigate phenomena like the moral credential effect [32,38], where a person who has chosen an environmentally friendly behavior in one context, may feel morally entitled to behave in less environmentally friendly fashion in another.

Collecting data on multiple environmental behavior 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 behavior (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 answers (in form of ticking checkboxes) to survey questions.

Furthermore, barcode scans could 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. The barcode scanning function could be linked to smart barcode scanner applications like GoodGuide or CodeCheck that allow the automatic identification of environmentally friendly products.

Data on electricity usage could be collected through taking a picture of the electric meter counter. 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 [15]. Users could then just correct potential errors (e.g. soya burger instead of a meat burger).

Furthermore, GPS records could be obtained to estimate travel distances and potentially to infer transportation modes [16]. It is, thus, suggested to combine sensing data (barcode scans, images, GPS records etc.) with ecological-momentary assessments data (survey answers), whereby the sensing data could be partly used to verify self-reported behavior. The challenge is to find a way to collect quickly and effectively sufficient and insightful data, without interfering too much with users' everyday life.

The captured data, that users would record through their day, could then be translated for instance into average CO_2 emissions, based on calculations provided for instance by [5]. This way, though various

behavior is recorded, all these behaviors are quantified in terms of environmental impact by a common measure. The translation of the collected data into a common quantifiable measure needs to be implemented within the data collection systems quasi real-time. This in turn would allow to produce some automatic simple descriptive analyses of the data in the background, which then would be represented through front-end visualizations in order to issue respective experimental interventions on a daily basis.

3.2. Multiple, bespoke experimental interventions

Another boundary to be pushed is on the experimental interventions. Presently, studies typically implement only a single experimental intervention and mostly it is a message-based intervention, where study participants are nudged to display a particular behavior. Other potentially interesting interventions have hardly 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 interventions are most effective in terms of real-life behavioral change. In the pilot study two interventions were implemented, but, in an actual study a set of various interventions should be tested.

Laboratory studies on public goods dilemmas can be very instructive in designing such interventions. Message-based interventions could be further developed through behavioral targeting, originally an online advertisement practice where online users are presented with advertisement based on their past online behavior [11]. Behavioral targeting is based in the nudge theory [48]. The "nudge" is any aspect "of the choice architecture that alters people's behavior in a predictable way without forbidding any options or significantly changing their economic incentives" (ibid.). In a behavioral targeting approach, individualized nudges, i.e. tailored messages could be sent to study participants' smartphones, proposing specific behavioral changes based on participants' past behavior. This has been implemented exemplary in the pilot study.

Another interesting intervention to be studied and implemented in the pilot study is social monitoring. Individuals in this treatment group will mutually monitor each other's behavior as captured by various environmental behavior scores and visualized through the smartphone application. It is assumed that people, who are aware of their behavior being monitored and who can compare their behavior to peer behavior, will tend to show socially desirable, i.e. environmentally friendly behavior [17,51]. This hypothesis is based in the social influence theory, which investigates the effects of compliance, conformity and competition [36].

Other field-experimental treatments that we did not have the means for to implement in the small pilot study, 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 [39]. A reputation-based intervention could be designed for instance by dividing the study participants into two competing groups, whereby both groups have the goal to collectively reduce their CO_2 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 would be ranked in each group based on their contribution to reducing CO_2 emissions and this rank would be visible to everyone in the group through the smartphone application. [54] for instance found in a large-scale field experiment that sufficiently high observability promotes cooperation in public good games much more effectively than financial incentives.

3.3. AI-based experimental interventions

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, e.g. nudging. Research moreover shows that in the long term, people are more likely to truly change their behavior towards the greater good, if they change their thinking; nudging for instance incites only superficial, non-lasting "first-order change" [53]. We should therefore explore ways to encourage behavioral change in a more emancipatory, empowering way, that is, empowering individuals in their capability to make informed decisions that are right for them and for the society overall.

There is research ongoing into designing algorithms and AI systems that do not violate people's autonomy and preserve their self-determination [44]. Building on this work we should design emancipatory interventions in future. Specifically, we suggest exploring the possibility of designing 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 CO_2 emissions), as implemented for instance in [34] and answer their questions (e.g. what are the alternatives). Additionally, and this would be the emancipatory intervention, the AI-based assistant could engage the users in a Socratic dialogue, encouraging them to think critically and make an autonomous and yet responsible decision, considering society's greater good and question their preferences and habits. First AI-based chat bots that engage users in Socratic dialogue to enhance critical thinking have been already built and studied [25]. This emancipatory intervention could be tested against a "non-emancipatory" intervention, which would for instance nudge the user to make a certain decision as described above.

3.4. Collecting additional data and data collection issues

Additionally to behavioural data and depending on the respective research question, data could be also collected on reasons for respective behaviors (e.g. why a certain transport mode was chosen), in order to understand how people rationalize their behavior. Similarly, if the influence of peers and other close contacts is of interest, data on who was involved in a given situation could be collected. An anonymised sociogram could be established for each study participant and stored within the app, study participants would than be able to link persons from the sociogram to a given environmental behaviour situation. This would allow to detect for instance whether some individuals display always a certain environmentally problematic behaviour if they are together with a specific person. Ultimately however, a parsimonious design with the aim to answer very specific research questions needs to be adopted to ensure the study participants are not overwhelmed and to comply with the EU GDPR (General Data Protection Regulation) guidelines. Hence, if certain queries are added others have to be removed.

Generally, collecting individual data requires careful consideration of privacy issues, as well as informational self-determination. We suggest collecting data anonymously. Moreover, the study participants should always have the complete control of their data and should at any point be able to decide what data they want to share. In the chapter below it will be described how privacy and informational self-determination was implemented in the pilot study.

4. Pilot study

To show the general feasibility of the above described approach, a pilot study was conducted over two weeks in June 2017 with 20 study participants. Two field-based interventions were tested to inspire cooperative, i.e., environmentally-friendly behavior: (1) behavioral 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 sufficiently large, representative sample should be aimed for in an actual study. Ethical approval has been obtained for the pilot study from the Faculty Research Ethics Committee at the University of Leeds (reference number: LTPOLIS-024).

4.1. Mobile phone application

The free EpiCollect 5 Mobile and Web Application (https://five.epicollect.net), developed by Imperial College London, was used for data collection purposes [1,2]. The platform allows 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. Work is ongoing on a bespoke smartphone application solution. This bespoke smartphone application automatizes many tasks that had to be performed by a human operator in the pilot study because of the limitations of EpiCollect 5. The bespoke app performs for instance simple statistical analysis of the collected data and produces graphs summarizing users' and group environmental performances (for the social monitoring intervention) that can be accessed through the application. Furthermore, the app automatically sends out nudge (for the behavioral targeting intervention) and reminder messages under certain conditions (e.g. user has reached a certain CO_2 emission level on some behavioral dimension, or user has not uploaded data by a certain time). This automatization allows for scalability, i.e. running the study on a large scale with many more study participants (see Supplementary Information S1 for further discussion on smartphone application, including ongoing work on a bespoke solution).

4.2. Data collection

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 a 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. Students were recruited through advertisement of the study and search for study participants in lectures at the department where the main author is teaching. Colleagues

as well as postgraduate students were recruited through University-based mailing lists. When recruiting study participants it was attempted to maintain the non-interference assumption, i.e. that the CO_2 emissions of one experimental group are not affected by the treatment in the other experimental group [21] e.g. through a spill-over effect between study participants, who are friends. This was done by recruiting students from different disciplines, graduation levels or courses.

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 CO_2 emissions for the specified activity based on [5]. This allowed to calculate average CO_2 emissions for each environmental behavior dimension and overall (see Supplementary Information S2.1 for further discussion of data collection, including questionnaire implemented in the app).

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 automatizing data collection, e.g. usage of accelerators for transport mode inference [16].

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 behavioral targeting group they would receive individualized messages giving advice on how they could reduce their CO_2 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 that visualized their own environmental performance from the previous day as well as the environmental performance of the others in the group via bar graphs. This happened in an anonymized way. Each study participant had a username (which strictly should not resemble the name or any other identification feature of the user) 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 5 pm (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, helped to maintain the excludability assumption, that is, the potential outcome of the experiment depends solely on what treatment the subject receives [21].

4.3. Discussion

The results of the pilot data analysis are discussed in the Supplementary Information S3, since the pilot study, given its various limitations (small sample, no control group, too short period for consistent
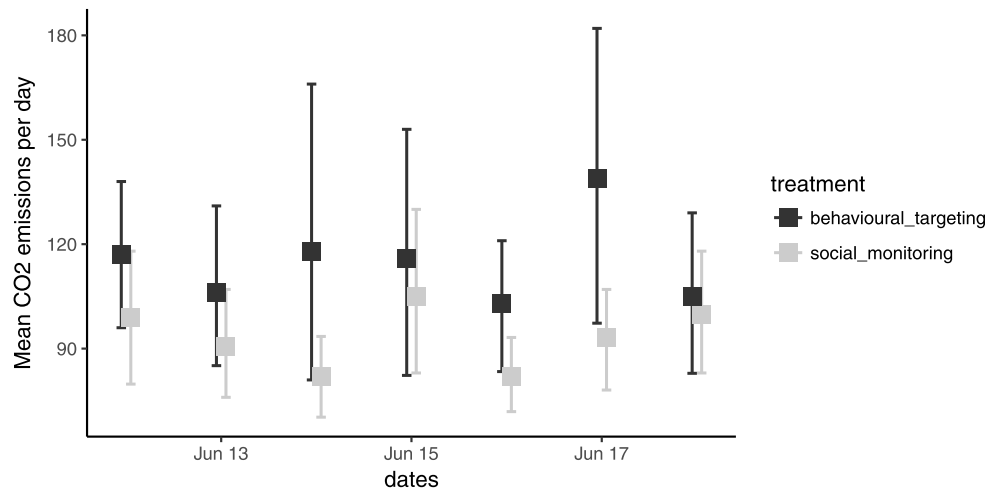
Fig. 1. Interaction plot shows the natural mean of each treatment*date combination along with the confidence interval of each mean with percentile method. Second week data used only, since no interventions implemented in the first week. Least Square Means for CO_2 emissions in the behavioral targeting group were estimated to be 118.02 (se: 8.61, 95%-CI: [97.12, 138.93]), for social monitoring 93.19 (se: 8.52, 95%-CI: [72.42, 113.96]). The difference is significant with t = 6.48 and $p < 0.01$.

behavioral change), does now allow to actually make conclusions about environmental behavior. Given these limitations many of the statistical results are inconclusive, nevertheless, the analyses of the pilot study data did produce some interesting results. For instance, comparing the two field-experimental interventions using the second week data seems to suggest that the social monitoring treatment had a somewhat greater positive (in terms of reducing CO_2 emissions) effect on the overall environmental behavior in comparison to the behavioral targeting treatment (see Fig. 1, and for further details Supplementary Information S3.3). This result shows that it could be worth investigating further the effects of different types of treatments in a full study, including a control-group and other treatments. Overall the data analyses 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.

There is much room for 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 among others valuable input for a better design 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 behavior. The problem of including more detailed queries however, is 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, for instance in terms of behavioral targeting messaging [14] and in terms of unobtrusiveness [21]. 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.

5. Conclusion

This paper has discussed how smartphones are increasingly used in studying human behavior. The study designs are getting increasingly sophisticated and so is the data collected through these devices.

This paper suggests that we can push the boundaries even further in order to be able to study more complex, multi-dimensional human behavior, which is not easily measurable and where various behaviors interact with each other. It was moreover argued that we can push the boundaries in terms of the experimental design of these studies, going beyond the often rather simple nudge-based messaging interventions. After all, nudging is highly controversial and potentially not even very effective in changing human behavior as discussed above. Smartphone-based studies allow to conduct field-experiments in an innovative way, where human behavior can be studied in real life. But, for these studies to become truly insightful and impactful we need to be able to test various interventions similar to those tested in laboratory experiments (incl. large online experiments) and beyond. This paper is making some suggestions of how we can move in this direction.

The paper focusses on environmental behavior as an example of social dilemma behavior. But sustainability is a complex and multi-faceted phenomenon associated with several issues. The moral credential effect was mentioned earlier and it necessitates studying different types of environmental behaviors and their interaction. Another effect, not mentioned earlier and not included explicitly in the pilot study is the so-called rebound effect, i.e. increased efficiency (and resulting savings from the increased efficiency) for instance in energy usage may lead to greater use of electricity or increases in other carbon-intensive behaviors. The rebound effect makes it also necessary to study different types of environmental behaviors simultaneously. And it shows how important it is to encourage deep-rooted, conscious behavioral change based on active decision making that prioritizes sustainability. Emancipatory interventions outlined in this paper are aiming at achieving such behavioral and cognitive change. But, sustainability is not only a problem to be dealt with at the individual level. It is not the intention of this study approach 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 or other interventions will not be sufficient to find a solution to the ecological crises humanity is facing. [22] 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 CO_2 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.

Supplementary data

Supplementary material is available at: http://dx.doi.org/10.3233/DS-180014.

Acknowledgements

References

[1] D.M. Aanensen, D.M. Huntley, E.J. Feil, F. al-Own and B.G. Spratt, EpiCollect: Linking smartphones to Web applications for epidemiology, ecology and community data collection, *PLoS ONE* **4** (2009), e6968. doi:10.1371/journal.pone.0006968.

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

[3] M. 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.

[4] 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.

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

[6] A. Biel, D. Elek, T. Gaerling and M. Gustaffson (eds), *New Issues and Paradigms in Research in Social Dilemmas*, Springer, 2008. ISBN 0387725954.

[7] 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, in: *Proceedings of the 8th International Conference on Ubiquitous Computing and Ambient Intelligence*, Belfast, UK, 2014, pp. 345–348. doi:10.1007/978-3-319-13102-3_57.

[8] 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.

[9] 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.

[10] 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.

[11] 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.

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

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

[14] 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.

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

[16] 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.

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

[18] 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.

[19] 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.

[20] 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/s12599-011-0201-5.

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

[22] 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.

[23] 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), 305–307. doi:10.1038/495305a.

[24] H. Gijzen, Development: Big data for a sustainable future, *Nature* **502** (2013), 38. doi:10.1038/502038d.

[25] 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.

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

[27] D. Griego, V. Buff, E. Hayoz, I. Moise and E. Pournaras, Sensing and mining urban qualities in smart cities, in: *Proceedings of the IEEE 31st International Conference on Advanced Information Networking and Applications (AINA)*, Taipei, Taiwan 2017, pp. 1004–1011. doi:10.1109/AINA.2017.14..

[28] 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.

[29] 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.

[30] 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.

[31] 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.

[32] 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.

[33] M.C.A. Klein, A. Manzoor and J.S. Mollee, Active2Together: A personalized m-health intervention to encourage physical activity, *Sensors* **17** (2017), 1436. doi:10.3390/s17061436.

[34] J. Klinglmayr, B. Bergmair, M.A. Klaffenböck, L.B. Hörmann and E. Pournaras, Sustainable consumerism via context-aware shopping, *International Journal of Distributed Systems and Technologies* **8** (2017), 54–72. doi:10.4018/IJDST.2017100104.

[35] 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.

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

[37] 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.

[38] 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.

[39] 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.

[40] 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.

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

[42] V. Pejovic and M. Musolesi, Anticipatory mobile computing for behaviour change interventions, in: *Proceedings of the ACM International Joint Conference on Pervasive and Ubiquitous Computing*, Seattle, USA, 2014, pp. 295–298. doi:10.1145/2638728.2641284.

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

[44] P. Pilgerstorfer and E. Pournaras, Self-adapting learning in decentralized combinatorial optimization – a design paradigm for sharing economies, in: *Proceedings of the IEEE/ACM 12th International Symposium on Software Engineering for Adaptive and Self-Managing Systems (SEAMS), Buenos Aires*, Argentina, 2017, pp. 54–64. doi:10.1109/SEAMS.2017.8.

[45] 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, in: *Proceedings of the 12th ACM International Conference on Ubiquitous Computing*, New York, USA, 2010, pp. 281–290. doi:10.1145/1864349.1864393.

[46] 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:10.1007/s10683-007-9190-0.

[47] 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.

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

[49] 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.

[50] 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.

[51] 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.

[52] 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, in: *Proceedings of the 2014 ACM International Joint Conference on Pervasive and Ubiquitous Computing*, New York, USA, 2014, pp. 3–14. doi:10.1145/2632048. 2632054.

[53] P. Watzlawick, J.H. Weakland and R. Fisch, *Change: Principles of Problem Formation and Problem Resolution*, Norton, New York, 1974. ISBN 0393707067.

[54] 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.

[55] 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. Available at: https://trid.trb.org/view.aspx?id=1371710.

[56] 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. Available at: https://www.utrc2.org/sites/default/files/pubs/Smart-Transportation-Choices-Mega-Cities-Smartphone-App.pdf.