Evaluation beyond Usability: Validating Sustainable HCI Research

Supplementary document: ingredients applied to example projects

PREFACE

This file is a supplement to the CHI 2018 paper "Evaluation beyond Usability: Validating Sustainable HCI Research" [1]. In this paper, we presented five ingredients that we believe can serve as guidance to identifying an appropriate evaluation process for an artefact in Sustainable HCI (SHCI) research. Our ingredients are intended as a starting point for a debate to the problem of how to evaluate the sustainable impact of HCI research, and therefore not considered to be final. However, in this document we aim to highlight how this first iteration of our five ingredients can provide insights into potential avenues for evaluation.

The examples below showcase potential ways to apply the ingredients to research projects. They do not serve as validation (or invalidation) of the ingredients itself; especially given that we applied the ingredients to our projects retrospectively as a thought exercise, and not during the course of the project itself. We aim to do this in the future, and for this purpose refer to our companion website (http://christianremy.com/evaluation) to which we welcome contributions from the community as well.

PROCESS

In the following, we elaborate on the concrete steps of how we applied the ingredients to our research, and caveats we observed. This was what worked for us – for other projects this process might differ.

Artefact

The first step is defining the artefact that is to be evaluated. This is typically the technological intervention that is created (e.g., an eco-feedback display or a sustainable electronic device), but it might also be a design idea (as highlighted in one of our examples). The definition of the artefact can be rather broad – however, it is important that it is a solution, and not the description of a problem. This is because otherwise we will encounter problems in the process of defining the other steps; already starting with the goal.

1. Goal

The goal of a research project usually is well-defined already; for the purpose of evaluation it might be beneficial to trying to be as specific as possible here. If we encounter problems with defining the goal, the artefact description itself might be the issue (is it really a solution, or at least a vision of a solution, for a sustainable problem?).

2. Mechanisms

Arguably, mechanisms are the most difficult part – but probably also the most important one. Mechanisms can be any effects caused by the artefact when deployed, on many different scales (from an individual's awareness to climate change). Brainstorming a broad variety of potential mechanisms along various dimensions (e.g., the three pillars of sustainability, the 17 Sustainable Development Goals of the UN, guidelines from previous SHCI research) can serve as a starting point to identify potential mechanisms. Do not exclude mechanisms that seem to be infeasible or unrealistic here; it can sometimes be helpful to point back to certain mechanisms in the discussion of the evaluation later, to showcase that one acknowledges and understands the complexity of related implications of the technological artefact.

3. Metrics

For each mechanism we ask: what is the measurable impact to observe whether or not the artefact has reached its goal? Sometimes there is no metric, or the metric is not measurable (e.g., the impact of a website on socio-political systems), which does not invalidate the mechanism – but it makes it out of scope for evaluation (see point 5).

4. Methods

Once we identified a mechanism and a metric, we need to discuss how we want to go about measuring it. This is usually connected to the metric itself and oftentimes we found ourselves considering the method already while thinking about potential metrics. Nevertheless, it is important to distinguish the two, as there can be several different methods to observe one metric, and it warrants a separate discussion to choose the one we feel most confident with.

5. Scope

The last step is to discuss all the different mechanisms/metrics/methods and choose the appropriate one for the final evaluation. This discussion depends on the project: sometimes there is only one feasible, sometimes there are multiple that can be combined, sometimes none of the mechanisms sound convincing (which might mean that we have to go back to identify other mechanisms).

EXAMPLE 1: Digital Support for Adaptive Thermal Comfort (Clear et al. 2014 [2])

Artefact	Augment HVAC to allow indoor temperature to fluctuate with the seasons, lowering the energy required. Provide digital tools (e.g. a web portal) to allow inhabitants to reflect on thermal experiences and adaptive measures.		
1. Goal	Raise awareness of and competency for body-focused adaptive measures such as clothing or hot drinks, to allow people to deal with a wider variety of temperatures indoors.		
2. Mechanism	A Reduced energy requirement of heating and cooling systems for the building.	B Because of interventions, use of clothing, hot drinks, and strategically timed exercise are worked into everyday practices.	C As the indoor temperatures vary more, participants will vary more in their adaptive measures, and experiment with new adaptive measures and timings.
3. Metric	Indoor room temperatures, energy expenditure (measured through fuel or electricity use), and other proxies such as temperature differential on radiator flow and return pipes.	Awareness of various adaptive measures. Characterisations of whether these kinds of things are normal or desirable.	Instances of employment adaptive measures, with respect to indoor room temperatures and outdoor conditions.
4. Method	Collect data on the metrics using sensors, and compare the differences before and after the intervention is deployed.	Interviews to discuss thermal comfort before and after the study. Surveys to assess sustained use of clothing, hot drinks and exercise for thermal comfort, periodically after the study has finished, and particularly after participants might have moved to other homes.	Interviews at the outset to establish views and competencies around clothing and other tactics for keeping warm/cool. Daily diary entries on usage of clothing, hot drinks, and exercise, before and after the deployment of the intervention. Indoor temperature sensors and records of outdoor temperature and precipitation.
5. Scope	Mechanism C is closest to the goal of the study, in its attempt to track engagement and experimentation with highly local, adaptive tactics for achieving thermal comfort. And crucially, how changes in the indoor environment and provision of digital support tools for reflection might alter competencies. Mechanism A would be a secondary priority, since it is nice to show how the change in indoor temperature is linked to energy savings. However, it is secondary because this link has already been extensively proven and modelled by the buildings research community. Mechanism B is important and certainly speaks to the long-term goals of interventions for adaptive thermal comfort, but is out of scope for this work, since there would be too many variables as participants move elsewhere: different indoor temperature control systems and different co-habitants.		

EXAMPLE 2: Energy demand of digital services on mobile devices (Widdicks et al. 2017 [3])

Artefact	Mobile phones and tablets, already in wide use in many parts of the world, allow people to watch, listen, communicate, and participate in online communities (social networking) through applications that run on the devices. These "apps" rely on (1) connectivity to the Internet via cellular and Wi-Fi; and (2) data centres and content distribution networks. This requires between 5-10% of global electricity, and is rising. Generally speaking, the more data traffic a service causes, the more energy intensive it is.			
1. Goal	Understand where and how the energy that goes into digital services supports everyday life.			
2. Mechanism	A The relative energy demand of different digital services in use.	B The value (utility and meaning) of digital services in the everyday life of the practitioner.	C Demand for services can continue across devices, and different services can be used concurrently with one another (e.g., instant messaging on a phone while streaming a film on TV).	D The continued use of devices is partially contingent on routines of charging and on software updates (however infrequent). People's approaches to these vary widely, but it impacts how services can continue to be used.
3. Metric	Energy is not directly measurable, but with care, the volume of data traffic can be used as a proxy for the energy demand of that service (http://dx.doi.org/10.1111/jiec.12630).	How often are different apps used? How does this support everyday practice in a practical way, and/or what meaning does it have?	Instances and durations of service use on different devices, concurrently, by the same person.	Times and durations of charging, and of software updates. Understandings of how people think about and organise charging and updates.
4. Method	A custom app running on the device monitors the volume of data traffic of different apps, over time.	A custom app which monitors frequency and duration of use of different apps throughout the day. Interviews to discuss use of the mobile device in daily life; supplemented by quantitative data on which apps are used and when.	Specialised software to capture times and durations of services on other devices in the home.	Custom software to monitor new software installs, and charging activity. Interviews to discuss approaches and attitudes to charging and updating.
5. Scope	Mechanisms A and B are crucial to tackling the goal of linking the digital services that support everyday practices, and the likely energy that results from this. A gives us a proxy (data traffic volume) for energy; and B allows us to understand whether this is meaningful or useful for what people do every day, or if it in fact has very little relevance. Mechanism D is certainly relevant, since charge level and software versions do affect when services can be called upon, how intensely that happens, and how much data traffic results. Charge level is not strictly speaking needed to accomplish the goal. We can get some sense for software updates by looking at traffic associated with the OS and app stores (which would happen via the methods in A). Thus, Mechanism D is related to the goal, but not directly in scope and thus not strictly necessary. Mechanism C addresses some larger and interesting questions about how digital services support activity in everyday life, but its methods would require significant effort to accomplish (e.g. custom software for a potentially huge variety of consumer devices). Note that both A and B assume that apps can be classified into different services, but this is not always true, for example web browsers support a variety of services and practices.			

EXAMPLE 3: Attachment in consumer electronics (Remy et al. 2015 [4])

Artefact	Designs that embody the attachment framework (Odom et al. 2009), i.e., foster a strong bond between owner and device, leading to long-term ownership and use. A group of product designers was asked to create designs of tablet computers promoting attachment; half the designers received the attachment framework whereas the other half acted as control group.			
1. Goal	Encourage long-term ownership and usage due to emotional attachment between owner and device, preventing premature disposal of consumer electronics and thus reducing obsolescence and e-waste.			
2. Mechanism	A The designs of the framework group are more likely to elicit emotional attachment and long-term use.	B The quality of the designs in general (including but not limited to attachment).	C Likelihood for any given design to be adopted in the consumer electronics domain, given economic pressure on manufacturers and retailers to sell devices.	D Ecological impact of the device considering projected long-term use.
3. Metric	Attachment is not a measurable metric itself, so one has to analyse the designs to argue for likelihood of a given design to embody the features of the attachment framework if realized in manufacturing.	Traditional design criteria - novelty, presentation, aesthetics, feasibility, etc.	Can't be measured and probably not even estimated, especially since the mechanism is contradicting the goal. However, one can justify/argue alternative income sources based on a design.	Requires combined metrics of both predicted duration of ownership as well as estimated environmental footprint.
4. Method	Analysis by sustainability/attachment experts to argue whether a design exhibits any of the attachment criteria.	Develop a list of typical design criteria and have a panel of experts assess the quality of the designs based on those criteria, either quantitatively (if numbers suffice) or qualitatively (to include justification and get a more detailed picture of how certain qualities are to be found in the designs).	Discussion of the potential effect of the design on the ecology/market of the device, and surrounding digital services or supporting physical peripheral devices or modules (depending on design). Linking the discussion to insights from marketing and/or economics and being upfront about caveats.	Predicted duration of ownership similar to metric A (sustainability/attachment expert's assessment of likelihood for long-term ownership); footprint could be estimated by metrics such as Life Cycle Assessment tools.
5. Scope	Mechanism A is the one that makes the least assumptions as it directly attempts to assess the impact of the framework on the resulting designs, but is subject to the ambiguity and subjectivity of attachment in design. Mechanism B takes into account that successful products need to exhibit "good design" in general - it is not enough to be just sustainable, and therefore is of critical importance. Especially due to the fact that the resulting artefacts are designs and not actual constructed and deployed technology products, mechanism B allows for assessing the "big picture". Mechanism C is one that is frequently discussed as device attachment is something that consumer electronics manufacturers might not support (as their preferred goal is rather brand attachment or attachment to the device ecology, rather than the actual device itself, allowing for replacement with newer versions of a device). However, it is important to acknowledge and discuss it. Mechanism D takes quite a leap of faith and is hampered due to a lack of data, in particular LCA data for consumer electronics, to allow for a holistic in-depth comparison of different devices - but this might change in the future.			

EXAMPLE 4: Environmental impact of last mile parcel deliveries (Bates et al. 2018 [5])

Artefact	Optimisation of delivery rounds that promote reduced vehicle mileage to reduce environmental impacts. Delivery round optimisation relies on a blend of optimisation software as well as planning and organisation by workers (drivers). Tools to help delivery drivers perform more sustainable parcel deliveries require hybrid optimisation (combining algorithms and human factors) in which variation in the driven mileage, number of stops, temporal and geographical overlap between rounds can impact air pollution, congestion and carbon emissions.			
1. Goal	Encourage a reduction in carbon emissions and pollution of parcel deliveries through changes in the vehicles used in parcel deliveries, changes in business models and through changing worker practices.			
2. Mechanism	A Increased flexibility in parking and on-foot decision making to encourage more walking.	B Increased efficiency of driver at the curbside by encouraging drivers to bundle parcels together that are geographically close and order appropriately onto van for fast retrieval at curbside once stopped.	C Increased knowledge in novice workers of geographic delivery patches.	D Optimisation of stopping locations due to reduced contribution to congestion and pollution.
3. Metric	Time spent on-foot during delivery round.	Time spent finding parcels in van and delivering once parked.	Increased knowledge, effectiveness, and performance of driver.	Number of deliveries made whilst stopped at a location within a 100 meter radius.
4. Method	Measuring the time spent on-foot per delivery round, comparing to before and after use of artefact.	Measure time spent retrieving parcels at curbside, comparing retrieval times before and after implementation of artefact.	Longitudinal interviews after deployment of tool to reflect on the driver's perceived knowledge, efficiency, and performance gain.	Longitudinal study to capture the number of deliveries completed from each stopping point within 100m radius before and after deployment of artefact to measure the change in driver practice within the context of each delivery round..
5. Scope	Mechanism D has the opportunity to be the most successful in terms of the goal as it encourages drivers to walk more (reducing mileage, congestion, pollution), encourages better utilisation of stopping places that are often time restricted in cities. Mechanism D has the opportunity to have the greatest impact towards the goal once scaled up across an entire delivery fleet, whereas the other mechanisms focus more on impacting the practices of individual drivers. Mechanism D also speaks to solutions that promote consolidation across parcel carriers at the curbside through the use of delivery vans as mobile delivery hubs.			

References

1. Remy, C., Bates, O., Dix, A., Thomas, V., Hazas, M., Friday, A., and Huang, E.M. (2018) Evaluation beyond Usability: Validating Sustainable HCI Research. In CHI' 18: Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems, Montreal, Québec, Canada (to appear).

2. Clear, A., Friday, A., Hazas, M., and Lord, C. (2014). Catch my drift?: Achieving comfort more sustainably in conventionally heated buildings. In Proceedings of the 2014 Conference on Designing Interactive Systems, DIS '14, pages 1015–1024, New York, NY, USA. ACM.

3. Widdicks, K., Bates, O., Hazas, M., Friday, A., and Beresford, A. R. (2017). Demand around the clock: Time use and data demand of mobile devices in everyday life. In Proceedings of the 2017 CHI Conference on Human Factors in Computing Systems, CHI '17, pages 5361–5372, New York, NY, USA. ACM.

4. Remy, C., Gegenbauer, S., and Huang, E. M. (2015). Bridging the theory-practice gap: Lessons and challenges of applying the attachment framework for sustainable hci design. In Proceedings of the 33rd Annual ACM Conference on Human Factors in Computing Systems, CHI '15, pages 1305–1314, New York, NY, USA. ACM.

5. Bates, O., Friday, A., Allen, J., Cherrett, T., McLeod, F., Bektas, T., Nguyen, T., Piecyk, M., Piotrowska, M., Wise, S., and Davies, N. (2018) Transforming Last-Mile Logistics: Opportunities for more Sustainable Deliveries. In CHI' 18: Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems, Montreal, Québec, Canada (to appear).