The Limits of Evaluating Sustainability

Christian Remy[1], Oliver Bates[2], Vanessa Thomas[3], Elaine M. Huang[1]

[1] University of Zurich, Switzerland
[2] School of Computing and Communications, Lancaster University, UK
[3] HighWire Centre for Doctoral Training, Lancaster University, UK
{remy, huang}@ifi.uzh.ch, {o.bates, v.thomas1}@lancaster.ac.uk

ABSTRACT

Designing technology with sustainability in mind is becoming more and more important, especially considering future scenarios of limited resources where the world's current lifestyle of wasteful consumption needs to change. But how can researchers believably argue that their solutions are indeed sustainable? How can consumers and technology users reliably acquire, understand, and apply information about environmental sustainability? Those questions are difficult to answer, especially in research domains where the impact on sustainability is not immediately measurable, such as sustainable HCI. The evaluation of sustainability is an ongoing problem that is often glossed over, but we believe the community needs to intensify its efforts to articulate its evaluation methods to other disciplines and external stakeholders. Even if those disciplines and stakeholders understand the importance of designing for sustainability, we need convincing arguments – such as validation through thorough evaluations – to showcase why a specific design solution works in the real world. In this paper, we analyze this problem by highlighting examples of sustainable HCI research in which evaluation of sustainability failed. We also look at previous research that sought to address this issue and discuss how their solutions can be generalized – and when they might fail. While we do not have the final answer, our intention is to start a discussion as to why sustainable HCI research is oftentimes not doing enough to justify the validity of its solutions. We close our paper by suggesting a few examples of what we believe to be potential ways to address those issues and take action to improve the evaluation of sustainability.

CCS CONCEPTS

• **General and reference → Evaluation** • **Human-centered computing → HCI design and evaluation methods** • **Social and professional topics → Sustainability.**

KEYWORDS

Evaluation; Sustainability; Sustainable HCI; SHCI; Sustainable Interaction Design.

1 INTRODUCTION

Within the HCI research community, scientific work is usually subject to a rigorous peer-review process, including when we publish papers at high-impact conferences or in journals. The review criteria differ from venue to venue, but usually include presentation, related work, originality, significance, and validity[1]. The first two – presentation and related work – are rather technical in nature. Originality often builds upon related work and is judged through arguments about why the proposed solutions fill a gap in the research landscape. For sustainability research, significance is usually clear because the scientific community is aware of the need for sustainable research. If a research project aims to create an impact for sustainability it is usually a significant contribution as long as the other criteria are fulfilled. But oftentimes the most difficult criterion is validity: How does one prove that a solution really addresses the identified problem at hand? How can we validate that the presented research reached its desired goals? In short: how do we measure success for sustainability?

To be able to answer those and other questions concerning the validity of research, a thorough evaluation of the proposed solution is necessary. In the field of Human-Computer Interaction (HCI), evaluation is an integral part of the design cycle [30] and an important activity that is included in basic HCI textbooks [e.g., 9, 41, 42]. However, there is no streamlined process or unified template that can be applied to every project in the same way; a novel artifact of technology requires an entirely different evaluation compared to a replication study. Sometimes an evaluation might even be harmful, e.g., for early and creative prototypes [15]. Sometimes presenting the empirical data of an ethnographic study is deemed sufficient to argue for validity [11]. Sustainable HCI (SHCI) research faces similar

---
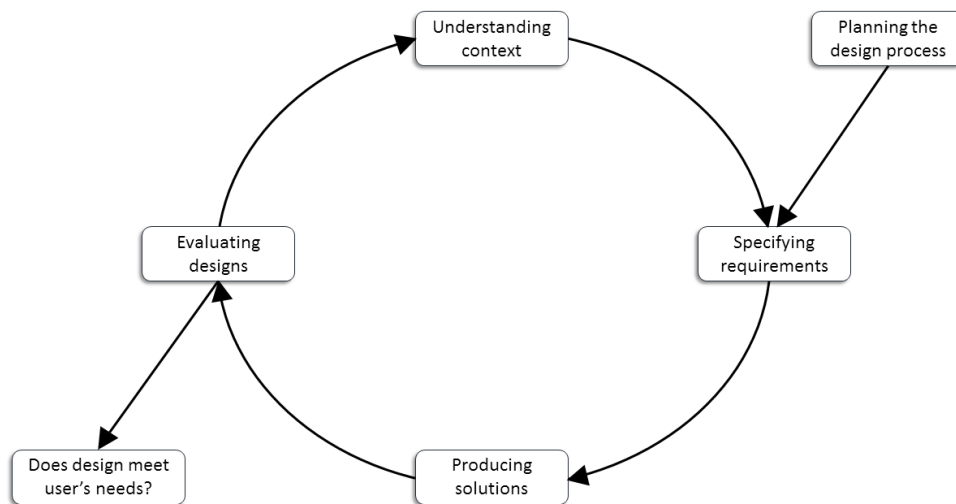
Figure 1: Interdependence of human-centered design activities (adapted from ISO 9241-210 [22]).

issues – finding the right way to evaluate a potential contribution is a difficult step, and oftentimes glossed over.

In this paper, we start by exploring reasons as to why evaluating SHCI is such a difficult endeavor and we reflect on previous discussions of this issue. As we have argued above, a thorough evaluation is mandatory to validate research, but it also serves as a means to promote research to practitioners outside the field. Furthermore, providing clear guidance on how to evaluate HCI research for sustainability can help other researchers contribute to sustainability and gain acceptance for their work in the SHCI community. We believe this is required to help grow the SHCI community and invite more research to address issues of sustainability. In the forthcoming pages, we will discuss examples of different strains of SHCI research and the challenges in evaluating those, as well as what needs to be done to address those challenges to arrive at a more rigorous evaluation.

We cannot present a generalizable solution for evaluating sustainability in HCI at this point – it would neither be feasible nor believable in a paper of this length, and it would only oversimplify a complex issue. Rather, we intend to start a discussion (or intensify existing discussions, where applicable) among the LIMITS community to acknowledge the issue, learn from mistakes or dead-ends of the past, and work towards a set of guidelines that can help researchers in the future. This is a critical issue, because if SHCI is limited in assessing the validity of its research, it is limited in communicating the value of its research, and therefore limited in creating an impact for sustainability. Our contribution in this paper is to discuss these limits and propose solutions for how to address them.

2 BACKGROUND: EVALUATION AND SUSTAINBLE HCI

2.1 Usability Evaluation in HCI

Whenever interaction designers create artifacts, regardless of whether those are digital or physical in nature, testing is an essential part in the design process. Dix et al. [9] summarize the three main goals of an evaluation as follows: *"to assess the extent of the system's functionality, to assess the effect of the interface on the user, and to identify any specific problems with the systems"* and Sharp et al. [41] note that *"[e]valuation is integral to the design process"*. How integral exactly can be determined if we consider the iterative design cycle usually employed in the human-centered design process (see Fig. 1, adapted from the ISO 9241-210 [22], as well as Sharp et al. [41]): without evaluation, there is no iteration and the design process breaks apart. In their seminal HCI textbook *Designing the User Interface*, Shneiderman et al. [42] stress the importance of evaluation by stating that *"[f]ailure to perform and document testing as well as not heeding the changes recommended from the testing process could lead to failed contract propoals [sic] or malpractice lawsuits from users where errors arise that may have been avoided"*. In short, there is no dispute within the HCI community that evaluation is an essential part of the discipline and not applying it rigorously can jeopardize the outcome of research.

There are limits to evaluating research. Obviously, not every contribution lends itself to a proper usability evaluation as mentioned in the aforementioned HCI textbooks. Submissions that focus on discussing theoretical concepts, reflect on the field and its methods, or are of a philosophical nature cannot be evaluated by traditional means such as usability guidelines or heuristics. In those cases, the validity of the contribution stems from factors such as the strength of the argument presented,

clarity in presenting the benefits for future research, and a thorough grounding in relevant literature.

However, even a piece of work that focuses on presenting a design artifact, such as a physical prototype or a web-based visualization, can offer a meaningful contribution to research without a typical evaluation. If the implementation is particularly creative and of unquestionable quality (e.g., by combining hardware and software in an ingenious way), the novelty and originality of the design solution might be sufficient to warrant deviating from typical evaluation practice. Most prominently, Greenberg and Buxton [15] argue that *"[u]sability evaluation, if wrongfully applied, can quash potentially valuable ideas early in the design process, incorrectly promote poor ideas, misdirect developers into solving minor vs. major problems, or ignore (or incorrectly suggest) how a design would be adopted and used in everyday practice."* Their prominent paper spearheaded a discussion that pervaded a major part of the HCI research domain, continued in prominent blogs accompanied by vivid discussions[2,3], and was followed by conferences de-emphasizing the importance of evaluation in favor of innovation and novelty (UIST 2010 in an email to the reviewers[4]). While this criticism is valid to date, it should be noted that Greenberg and Buxton close by stating that a traditional usability evaluation is the best method *"in many, but not all cases"* and *"in all cases a combination of methods – from empirical to non-empirical to reflective – will likely help to triangulate and enrich the discussion of a system's validity"* [15]. Many of the non-empirical methods they propose (*"design critiques, design alternatives, case studies, cultural probes, reflection, design rationale"* [15]) have since become the de-facto standards within the HCI community.

2.2 SHCI and Evaluation

SHCI emerged as a subfield of HCI at the CHI conference in 2007 [5, 29] and therefore saw itself subject to the same rigor in evaluating its research outcomes. As an emerging, young research area, many projects initially fell under the umbrella of innovative design artifacts. This is not to say that any of those early SHCI works lacked evaluation–quite the contrary. Breaking into an unclaimed field and touching new ground comes with other challenges, such as having to argue for relevance or appropriate context of the conducted research. However, as the SHCI community started looking back at the plethora of research projects it had created in a relatively short timeframe [8], more critical voices appeared to question the impact achieved by the SHCI community and some suggested different approaches [e.g., 6, 10, 14].

The field had adapted the standards of HCI, but also started to emphasize the need for an additional metric of measuring contribution: sustainable impact. A newly developed system must adhere to the traditional evaluation of systems in HCI research as well as prove that it achieves its goals towards

SIGCHI "sustainability" publications
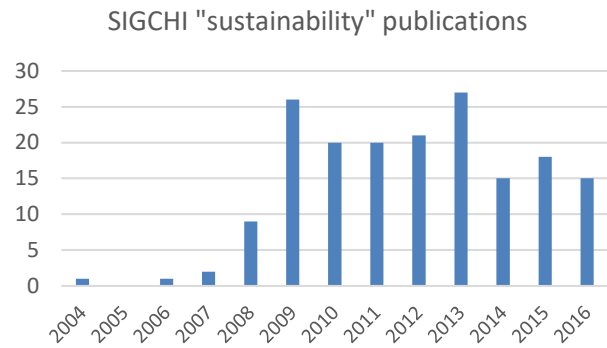
Figure 2: Publications at the SIGCHI conference series with author keyword "sustainability", based on an ACM Digital Library [1] search.

sustainability. However, there are no standardized metrics for assessing sustainable impact, and there is not even a clear definition of sustainability; recent workshops [12, 26] argued for the UN's Sustainable Development Goals [40] as a means of orienting SHCI research within the real world, but these have yet to be adopted by the broader SHCI community.

In terms of evaluation, the community has not made any significant inroads over the past ten years. Dillahunt et al. [7] discussed a framework to assess environmental sustainability, developed with the help of sustainability experts, which comes as a checklist of several sustainability criteria (e.g., *"Uses alternative energy"*, *"All materials can be replaced"*, *"All materials are reusable"*, *"Device is recyclable"*). Silberman and Tomlinson [44] suggest that SHCI *"could become more relevant by developing evaluations that link to understandings of sustainability beyond HCI"* and describe three tools for sustainability evaluation: principles, heuristics, and indices. None of the proposals gained traction, and four years later the community even agreed that a unified evaluation framework is unrealistic. At an SHCI workshop [43], most participants *"rejected the idea that [they] could devise a single interpretation of sustainability to orient and evaluate all future SHCI research"*. Rather, they concluded that SHCI projects should define their own goals and metrics, depending on the specific case, and consider criteria from outside of HCI as well.

The history of usability evaluation in HCI, including the criticism and reorientation that emphasized non-empirical evaluation methods, has taught us that there is no one-size-fits-all solution. After decades of research and the emergence of multiple usability heuristics, guidelines, and evaluation frameworks, HCI is still a field in motion that is evolving and considering new forms of assessing its contributions' values. Therefore, the SHCI community is likely taking the right step in not prescribing any strict rules for evaluation or prescribing any evaluation frameworks and heuristics; also in light of previous failed efforts to do so.

However, we argue that this freedom has become an obstacle: in many (if not most) cases, evaluating the sustainable impact is still a requirement to gain acceptance from the SHCI community

– but how are new researchers able to enter the field without any guidance whatsoever? In addition to the usual pressure of evaluating contributions by traditional HCI standards, one must conduct an additional evaluation for sustainable impact. This includes defining what sustainability means for the specific project, articulating the goals one wants to achieve, surveying fields outside of HCI for suitable metrics (e.g., social sustainability or material science), developing an entirely new evaluation method, and conducting said evaluation. While we agree that there are advantages to not prescribing a concrete process for evaluating sustainability, we believe that the current lack of guidance and clarity within the SHCI community might be contributing to the decline in sustainability-related publications at the SIGCHI conferences (see Fig. 2, [1]).

In the following, we will highlight examples from SHCI research – separated into the two different branches that divide the field thematically – to showcase the difficulty in evaluating sustainability. The purpose of those examples is twofold: first, we point out the limits in evaluating SHCI research and assessing sustainable impact; second, we discuss those examples in the discussion section and aim to start a conversation for potential solutions to the problem of evaluating sustainability.

3 THE LIMITS OF EVALUATING SHCI

SHCI research can roughly be divided into two different approaches: *sustainability through design* and *sustainability in design* [29]. *Sustainability through design* aims to develop technology that has an impact on sustainability through people's lifestyles, e.g., by visualizations that raise awareness or applications that promote behavior change. This line of research is often referred to as eco-feedback technology [14] or persuasive technology [6]. *Sustainability in design* is about developing technology that is sustainable regardless of use, e.g., by choosing recyclable material or enabling repair of a device. While sometimes used as synonym for SHCI, Blevis's initial concept of sustainable interaction design (SID) [5] is rather concerned with this direct approach to sustainability [38].

Both branches of SHCI have seen a sizeable amount of research in the past, however, they differ significantly in their goals, methods, and outcome. Therefore, it is imperative to discuss the difficulty of evaluation individually for each of those.

3.1 Sustainability through Design

Since the goal of *sustainability through design* is to affect the lifestyle of people who use said technology, the measure of success goes beyond that of traditional HCI solutions. If the technology holds up to the most rigorous usability evaluation but shows no effect on people's lifestyle, it has failed to achieve an impact for sustainability; or as Fogg [13] points out: *"[d]esigning for persuasion is harder than designing for usability"*. He recommends to test early (and often), a suggestion echoed by all HCI textbooks, and defines the goal as *"create an intervention that succeeds in helping the target audience to adopt a very simple target behavior that can be measured"*. However, in a comprehensive survey, Froehlich et al. state that *"few HCI eco-*

feedback have even attempted to measure behavior change" [14], and other SHCI scholars [e.g., 6, 8, 43] discussed the difficulty of measuring the impact of sustainability through design. What needs to be considered is the complexity that encompasses not only technology acceptance, classical usability, and measurable effects on the consumer's lifestyle, but also social contexts, environmental factors, and a myriad of additional variables – for which SHCI designers often lack the required knowledge and skills to assess those in proper scientific rigor.

Therefore, SHCI research oftentimes does not set its goal to change behavior, but rather to raise awareness. This acknowledges that behavior change is a process that develops over time, and it is separated into different stages. For example, the transtheoretical model [19, 35] comprises five stages (precontemplation, contemplation, preparation, action, maintenance), of which the actual behavior change takes place in the fourth stage (cf. [23] for other models). This does not alleviate the problem of evaluation, it merely transforms it: instead of measuring behavior change, one needs to measure raised awareness. Therefore, a common approach for persuasive technology in SHCI is to provide information (e.g., through visualizing environmental data) and rely on self-reported participant data or interviews to verify the information transfer. Knowles et al. criticize this as an undesirable solution and call providing information an anti-pattern: *"The implicit assumption of these designs is that greater awareness of their consumption will inspire users to change their behavior"* [25]. However, Knowles et al. also do not advocate going back to Fogg's initial evaluation of measuring behavior change, cautioning against rebound effects and other neglected contextual factors that are not being captured.

Ultimately, SHCI maneuvered itself into a difficult spot: the community demands a scientifically rigorous sustainability evaluation of any presented solution. At the same time, SHCI has a rich history of designs and evaluations that did not work – arguably more negative than positive examples as we have heard from fellow researchers and have experienced ourselves in the review process (both as authors and reviewers). What are potential solutions? How does one evaluate sustainability through design? SHCI researchers have discussed alternative ways of assessing the impact of persuasive technology, and we will list some of those here:

3.1.1 Large-scale deployments. Comparing studies in SHCI to psychology, Froehlich et al. [14] note that the sample size of studies in SHCI is remarkably smaller (11 vs. 210 participants on average, respectively). This is not necessarily a fair comparison – psychological studies are often controlled, quantitative experiments, whereas SHCI researchers seem to prefer early prototype tests in qualitative settings. Also, scaling up studies is likely to introduce additional problems as it is at odds with limited resources and time available to researchers, hardly works for low-maturity prototypes, and does impose even more rules on clearly defining the metrics of evaluation.

3.1.2 Long-term studies. Researchers who want to measure the impact on participants objectively should aim for a longer timeframe; in their survey of persuasive technology,

Brynjarsdottir et al. [6] consider only one study with a duration of three months as long-term study. The transtheoretical model suggests that behavior needs roughly six months to settle in [19], and to also pay justice to the fifth stage of "maintenance" with potential relapse, a one-year timeframe is advised. Time limits on researchers' projects often prohibit such long-term evaluations, as contracts, grants, or doctoral programs are difficult to unite with such commitments.

3.1.3 Participatory design. Fogg [13] recommends to test and iterate designs early and often, and HCI textbooks also emphasize that it is advisable to evaluate designs throughout, instead of just adding an evaluation at the end of the process [9, 41, 42]. Participatory design ensures that evaluation occurs throughout design processes, and SHCI researchers have previously recommended to include the user into the design process [6].

3.1.4 Different models. An evaluation measures the effect of a design artifact against the design goal and requirements (cf. Figure 1). If researchers struggle with the evaluation, it might sometimes be a symptom of not clearly enough defining the goal beforehand. Choosing a different background, such as He et al. [19] did with the transtheoretical model of behavior change, could potentially address this [14]. However, the number of existing models of behavior change is limited [23] and fully understanding and implementing them introduces new obstacles to the process (limited time and resources). Suggested alterations to the evaluation process are to focus on practices of the users [6, 33] or users' reflections of provided information [25] as a middle ground for the overambitious goal "behavior change" and the superficial approach to simply "provide information".

3.2 Sustainability in Design

The contributions regarding *sustainability in design* in the field of SHCI are more theoretical and offer fewer design artifacts than the contributions found in much *sustainability through design* research. Blevis's rubric [5] for understanding and assessing the material effects of interaction design was pivotal for the field of SHCI. Several studies were conducted to further investigate people's practices relevant to SID [e.g., 17, 20, 21, 28, 31, 32] and multiple frameworks and guidelines deepened our understanding of SID by focusing on specific themes, such as re-use [24], attachment [31], or cloud computing [34].

However, there are few examples of design artifacts from SHCI research that seek to apply those frameworks to practice, and even fewer that attempt to evaluate them. Two exceptions are design exercises with practitioners who created solutions by implementing theoretical frameworks: slow design [16] and attachment [36]. The result of the slow design exercise was a mock-up prototype, which was being evaluated by six workshop participants who reflected on the imagined use of the prototype in their everyday life. The second example of applying the attachment framework to design practice was conducted as a comparative study, and the resulting designs were evaluated by design experts for traditional design criteria along with attachment. Besides the apparent differences in study design and evaluation (one prototype vs. multiple design sketches; reflection of potential scenarios vs. assessing inherent design qualities), the studies have a few things in common. Both evaluate SID early in the design process and at the start of a potential product's lifecycle; both recruit external evaluators for an objective assessment; and both projects assess the effect of the framework qualitatively rather than focusing on measurable, quantitative metrics.

Evaluating SID is difficult; so difficult, in fact, that the evaluation itself took more time than the rest of the exercise in the attachment study [36]. Reviewers of the paper considered the evaluation process a major contribution. This is an issue similar to persuasive technology, for which SHCI asks the researchers to create their own metrics rather than providing a template for evaluation. By expressing interest in applying SID to the design process, SHCI puts the burden of creating an evaluation entirely on the researcher; but not every researcher has the time, expertise, or desire to develop new evaluation methods. Blevis rightly argues that *"sustainability can and should be a central focus of interaction design"* [5] – but in order to achieve this, SHCI needs to provide guidance for how to evaluate this shift in focus. Without evaluating the effect that adding sustainability to interaction design has, there is no proof for the validity of a design solution.

Evaluating SID is also a matter of feasibility. When implementing the slow design or attachment frameworks into a product's design process, one might argue that the only real measure of success would be to observe the objects in practice, similar to real-world deployments of persuasive technology. However, designing, building, and distributing products, and then being able to evaluate their use years later is far beyond the limits of most feasible research projects. Therefore, an evaluation needs to be employed at the early stages of design – which is in line with the idea of HCI's iterative design cycle. It also has an added benefit: mistakes can be discovered early in the process when design decisions are still reversible. The drawback is that those early evaluations come with a lot of ambiguity [36, 44].

Due to the theoretical nature of SID and the limited examples of actual evaluations, the list of potential solutions for this issue is of rather anecdotal nature. Nevertheless, we will highlight themes that have been mentioned within the community or came up during our own struggles with evaluating SID in practice:

3.2.1 Evaluate prototypes and ideas. As highlighted in the example of attachment, it is often not feasible to evaluate SID in real-world scenarios with design artifacts of high maturity; this is partially due to the constraints on researchers' time and resources, but also due to the limited time left to save the environment before the damage from our non-sustainable lifestyles becomes irreversible. Therefore, SHCI research needs to be accepting of early prototypes or even rough sketches of ideas how SID could be applied to practice and what those solutions might look like. This is not to recommend neglecting scientific rigor in evaluating such applications; however, the community needs to work towards accepted standards for what constitutes a successful application and be mindful of the difficulties in designing and evaluating those.

3.2.2 Evaluate the process, not the product. Applying design research theory to design practice is difficult; it is a well-known issue that is often referred to as the theory-practice gap. Figuring out how to address the theory-practice gap [37, 39] has potential to be a valuable contribution to SHCI (as well as HCI in general). But addressing the theory-practice gap remains a challenge because there is no standard metric for measuring the transfer of knowledge from one domain to another. While the theory-practice gap has been a known problem in HCI for several decades, the urgency of combating environmental issues does not allow for SHCI to wait for a solution. The community needs to find ways to give researchers a chance to argue for success of their process of sustainable interaction design instead of waiting for its outcome to be evaluated.

3.2.3 Outsourcing evaluation. In our two highlighted examples of applying SID to design practice [16, 36], the researchers did not conduct the evaluation themselves, but recruited external evaluators. This might generally be good practice to maintain objectivity and enables SHCI to recruit experts who bring in additional expertise. However, it adds to the difficulty of evaluation, as it requires time and resources (compensation for the experts), but most importantly it requires a common understanding of the goals. Every discipline has their own terminology and jargon, and SHCI is no different; establishing a lingua franca for the evaluation of SID by externals might help to streamline this process.

3.2.4 Resource assessment. In cases where it is applicable (i.e., when the impact of a designed SID artifact is measurable), other disciplines might help SHCI to address resource assessment. For example, if a solution proposes the use of different material (hardware design) or argues for a lower environmental impact of an algorithm (software design), one metric to evaluate success can be to calculate the resources saved. The most prominent example to achieve this is life cycle assessment (LCA), which offers a holistic overview of a product's environmental impact based on a variety of different metrics. While LCA is a work-in-progress and therefore has limitations on its own, it should at least be considered as an additional metric if applicable. There have already been a few early attempts at blending approaches from LCA with design to consider the environmental impact of digital technology in practices [3, 4] and home energy intervention studies [2], as well as using methods for mitigating the growing impact of data demand generated by mobile digital technology [18, 27].

4 DISCUSSION

We have looked at the general process of evaluating design artifacts in HCI, the difficulty of evaluating SHCI specifically, and provided some pointers for potential solutions. As mentioned before, in particular the research contributions surrounding SID are often of theoretical nature and therefore not subject of our discussion. There is also a great deal of research studying people and technology, and those studies are not subject to a traditional usability evaluation either (see Dourish's concerns about implications for design [11] for a discussion

about how to present the results of ethnographic studies). While we acknowledge that those papers are excluded from our discussion and provide invaluable insights for the field of SHCI, we believe the balance is off. There needs to be more applications of theoretical insights to practice, otherwise the theoretical discussions will stay exactly that – theoretical – and never have an impact on sustainability issues in the real world. We believe that the limits to evaluating sustainability—whether they are limits perceived by new researchers seeking to break into sustainable research or limits observed by long-term members of the field—pose a threat to SHCI.

The SHCI community stressed that there can be no evaluation that fits all research. They asked researchers to define *"design-specific sustainability goals and metrics on a project-by-project basis"*, and include criteria from *"the communities within which they work"* [43]. This is a laudable approach and we should cherish it, as it promotes diversity of thought and pays justice to the complexity of our environment. However, it can also backfire, as researchers entering the field might not be familiar with SHCI's processes and expectations; also, they might not have the expertise or willingness to define their own goals and metrics. But most importantly, they might be driven away from SHCI by focusing on a different own community's goals, as the overwhelming majority of HCI communities have not included sustainability in their processes yet. Asking them to adhere to their community's standards for evaluation is equivalent to asking them to neglect sustainability.

Therefore, SHCI needs to provide at least rough guidance for the overarching goals of the field. Recent SHCI workshops have started to do this [12, 26] by pointing to the United Nation's Sustainable Development Goals [40] as means of guidance. However, similar to the seminal SID rubric, it can only be the starting point for developing specific goals, metrics, and processes for evaluation. It is also important for the community to acknowledge that establishing a goal and defining metrics does not eliminate the process of evaluation. A goal defines the desirable endpoint of a project, and metrics enable the assessment for concluding whether the goal has been reached or not (and by how much). But evaluation is the process that connects everything by interpreting the solution in light of the previously defined metrics.

One approach for addressing the problem of evaluating sustainability is to continue the work that SHCI excels at: learning from other disciplines by understanding and adapting their methods. We already mentioned LCA as a potential means to assess the measurable impact of SID solutions. Another example is the BELIV workshop series[5], a biennial event that discusses novel evaluation methods for visualization, which might provide helpful pointers for eco-feedback technology if extended by sustainable criteria. The process of bridging disciplines can be difficult, but SHCI has shown its capabilities to do so by incorporating numerous external aspects into its research. Through this, SHCI has created various theoretical

---

[5] http://beliv.cs.univie.ac.at/

frameworks. It is time to shift our attention away from drawing theoretical lessons and towards the evaluation of practice.

Improving the process of evaluating sustainability for the purposes of our research might also help the field in different ways. It enables SHCI to argue for the validity of its findings when communicating to other stakeholders, such as product designers [36] or policymakers [45]. For certain aspects of SHCI research it is even essential to be able to evaluate sustainability: How can users of eco-feedback technology be expected to evaluate their own lifestyles against the provided information if the researchers are not able to do so themselves? How can one teach sustainability without a holistic understanding thereof?

We have highlighted the two different branches of SHCI, and believe the problem of evaluation needs to be solved for both – but separately. Even within those two areas, the most suitable evaluation method depends on many different factors, including the maturity of the proposed design solution. While fully developed prototypes can be evaluated in real-world deployments, low-maturity concepts should rather be subject to evaluation of domain experts who understand and can look beyond the level of abstraction. Therefore, we urge the community to not confuse metrics with maturity, and instead of choosing the evaluation method based on the available measurement or sustainable goals to focus on what is most appropriate given the state of the solution's development. While SID's rubric [5] or the Sustainable Development Goals [40] are helpful for establishing research goals, they are not complete solutions to evaluation, and they are unlikely to be the best labels to categorize different evaluation methods for SHCI.

5   CONCLUSION

In this paper, we argue that SHCI research is often glossing over the evaluation of its results. This has led to a situation in which both new researchers coming into the field as well as long-time members of the community lack guidance on how to evaluate their results. We believe that this not only hurts the validity of research conducted in SHCI, but also threatens its credibility and standing within the larger community of HCI research, and is alienating rather than attracting more research to consider orienting their work towards important sustainable issues.

Following our analysis of evaluation in HCI in general and SHCI in particular, we outlined several pointers which can help addressing this issue. Although we do not have a solution for how to evaluate all future SHCI research, we hope our arguments are perceived as constructive criticism to solve a problem that we believe is threatening the core of the SHCI community. Our intention is to start a discussion within SHCI and, in a best-case scenario, arrive at a community-based repository for evaluating SHCI research.

REFERENCES

[1]   ACM Digital Library: 2017. http://dl.acm.org/. Accessed: 2016-10-17.

[2]   Bates, O. and Hazas, M. 2013. Exploring the Hidden Impacts of HomeSys: Energy and Emissions of Home Sensing and Automation. Proceedings of the 2013 ACM Conference on Pervasive and Ubiquitous Computing Adjunct Publication (New York, NY, USA, 2013), 809–814.

[3]   Bates, O., Hazas, M., Friday, A., Morley, J. and Clear, A.K. 2014. Towards an Holistic View of the Energy and Environmental Impacts of Domestic Media and IT. Proceedings of the 32Nd Annual ACM Conference on Human Factors in Computing Systems (New York, NY, USA, 2014), 1173–1182.

[4]   Bates, O., Lord, C., Knowles, B., Friday, A., Clear, A. and Hazas, M. 2015. Exploring (un)sustainable growth of digital technologies in the home. (Copenhagen, Denmark, 2015).

[5]   Blevis, E. 2007. Sustainable interaction design: invention & disposal, renewal & reuse. Proceedings of the SIGCHI Conference on Human Factors in Computing Systems (New York, NY, USA, 2007), 503–512.

[6]   Brynjarsdottir, H., Håkansson, M., Pierce, J., Baumer, E., DiSalvo, C. and Sengers, P. 2012. Sustainably unpersuaded: how persuasion narrows our vision of sustainability. Proceedings of the 2012 ACM annual conference on Human Factors in Computing Systems (New York, NY, USA, 2012), 947–956.

[7]   Dillahunt, T., Mankoff, J. and Forlizzi, J. 2010. A proposed framework for assessing environmental sustainability in the HCI community. Examining Appropriation, Re-Use, and Maintenance of Sustainability workshop at CHI 2010 (2010).

[8]   DiSalvo, C., Sengers, P. and Brynjarsdóttir, H. 2010. Mapping the landscape of sustainable HCI. Proceedings of the SIGCHI Conference on Human Factors in Computing Systems (New York, NY, USA, 2010), 1975–1984.

[9]   Dix, A.J., Finlay, J.E., Abowd, G.D., Beale, R. and Finley, J.E. 1998. Human-Computer Interaction. Prentice Hall.

[10]   Dourish, P. 2010. HCI and environmental sustainability: the politics of design and the design of politics. Proceedings of the 8th ACM Conference on Designing Interactive Systems (New York, NY, USA, 2010), 1–10.

[11]   Dourish, P. 2006. Implications for Design. Proceedings of the SIGCHI Conference on Human Factors in Computing Systems (New York, NY, USA, 2006), 541–550.

[12]   Eriksson, E., Pargman, D., Bates, O., Normark, M., Gulliksen, J., Anneroth, M. and Berndtsson, J. 2016. HCI and UN's Sustainable Development Goals: Responsibilities, Barriers and Opportunities. Proceedings of the 9th Nordic Conference on Human-Computer Interaction (New York, NY, USA, 2016), 140:1–140:2.

[13]   Fogg, B. 2009. Creating Persuasive Technologies: An Eight-step Design Process. Proceedings of the 4th International Conference on Persuasive Technology (New York, NY, USA, 2009), 44:1–44:6.

[14]   Froehlich, J., Findlater, L. and Landay, J. 2010. The design of eco-feedback technology. Proceedings of the SIGCHI Conference on Human Factors in Computing Systems (New York, NY, USA, 2010), 1999–2008.

[15]   Greenberg, S. and Buxton, B. 2008. Usability Evaluation Considered Harmful (Some of the Time). Proceedings of the SIGCHI Conference on Human Factors in Computing Systems (New York, NY, USA, 2008), 111–120.

[16]   Grosse-Hering, B., Mason, J., Aliakseyeu, D., Bakker, C. and Desmet, P. 2013. Slow Design for Meaningful Interactions. Proceedings of the SIGCHI Conference on Human Factors in Computing Systems (New York, NY, USA, 2013), 3431–3440.

[17]   Hanks, K., Odom, W., Roedl, D. and Blevis, E. 2008. Sustainable millennials: attitudes towards sustainability and the material effects of interactive

technologies. Proceedings of the SIGCHI Conference on Human Factors in Computing Systems (New York, NY, USA, 2008), 333–342.

[18] Hazas, M., Morley, J., Bates, O. and Friday, A. 2016. Are There Limits to Growth in Data Traffic?: On Time Use, Data Generation and Speed. Proceedings of the Second Workshop on Computing Within Limits (New York, NY, USA, 2016), 14:1–14:5.

[19] He, H.A., Greenberg, S. and Huang, E.M. 2010. One size does not fit all: applying the transtheoretical model to energy feedback technology design. Proceedings of the SIGCHI Conference on Human Factors in Computing Systems (New York, NY, USA, 2010), 927–936.

[20] Huang, E.M. and Truong, K.N. 2008. Breaking the disposable technology paradigm: opportunities for sustainable interaction design for mobile phones. Proceedings of the SIGCHI Conference on Human Factors in Computing Systems (New York, NY, USA, 2008), 323–332.

[21] Huh, J., Nam, K. and Sharma, N. 2010. Finding the lost treasure: understanding reuse of used computing devices. Proceedings of the SIGCHI Conference on Human Factors in Computing Systems (New York, NY, USA, 2010), 1875–1878.

[22] ISO 9241-210:2010 - Ergonomics of human-system interaction -- Part 210: Human-centred design for interactive systems: https://www.iso.org/standard/52075.html. Accessed: 2017-03-10.

[23] Jackson, T. 2005. Motivating Sustainable Consumption. Centre for Environmental Strategies, University of Surrey, UK.

[24] Kim, S. and Paulos, E. 2011. Practices in the creative reuse of e-waste. Proceedings of the SIGCHI Conference on Human Factors in Computing Systems (New York, NY, USA, 2011), 2395–2404.

[25] Knowles, B., Blair, L., Walker, S., Coulton, P., Thomas, L. and Mullagh, L. 2014. Patterns of Persuasion for Sustainability. Proceedings of the 2014 Conference on Designing Interactive Systems (New York, NY, USA, 2014), 1035–1044.

[26] Knowles, B., Clear, A.K., Mann, S., Blevis, E. and H\a akansson, M. 2016. Design Patterns, Principles, and Strategies for Sustainable HCI. Proceedings of the 2016 CHI Conference Extended Abstracts on Human Factors in Computing Systems (New York, NY, USA, 2016), 3581–3588.

[27] Lord, C., Hazas, M., Clear, A.K., Bates, O., Whittam, R., Morley, J. and Friday, A. 2015. Demand in My Pocket: Mobile Devices and the Data Connectivity Marshalled in Support of Everyday Practice. Proceedings of the 33rd Annual ACM Conference on Human Factors in Computing Systems (New York, NY, USA, 2015), 2729–2738.

[28] Maestri, L. and Wakkary, R. 2011. Understanding repair as a creative process of everyday design. Proceedings of the 8th ACM conference on Creativity and cognition (New York, NY, USA, 2011), 81–90.

[29] Mankoff, J.C., Blevis, E., Borning, A., Friedman, B., Fussell, S.R., Hasbrouck, J., Woodruff, A. and Sengers, P. 2007. Environmental sustainability and interaction. CHI '07 Extended Abstracts on Human Factors in Computing Systems (New York, NY, USA, 2007), 2121–2124.

[30] Nielsen, J. 1994. Usability engineering. Morgan Kaufmann Publishers.

[31] Odom, W., Pierce, J., Stolterman, E. and Blevis, E. 2009. Understanding why we preserve some things and discard others in the context of interaction design. Proceedings of the SIGCHI Conference on Human Factors in Computing Systems (New York, NY, USA, 2009), 1053–1062.

[32] Pierce, J. and Paulos, E. 2011. Second-hand interactions: investigating reacquisition and dispossession practices around domestic objects. Proceedings of the SIGCHI Conference on Human Factors in Computing Systems (New York, NY, USA, 2011), 2385–2394.

[33] Pierce, J., Strengers, Y., Sengers, P. and Bødker, S. 2013. Introduction to the Special Issue on Practice-oriented Approaches to Sustainable HCI. ACM Trans. Comput.-Hum. Interact. 20, 4 (2013), 20:1–20:8.

[34] Preist, C., Schien, D. and Blevis, E. 2016. Understanding and Mitigating the Effects of Device and Cloud Service Design Decisions on the Environmental Footprint of Digital Infrastructure. Proceedings of the 2016 CHI Conference on Human Factors in Computing Systems (New York, NY, USA, 2016), 1324–1337.

[35] Prochaska, J.O. and Velicer, W.F. 1997. The transtheoretical model of health behavior change. American journal of health promotion: AJHP. 12, 1 (Oct. 1997), 38–48.

[36] 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. Proceedings of the 33rd Annual ACM Conference on Human Factors in Computing Systems (New York, NY, USA, 2015), 1305–1314.

[37] Roedl, D.J. and Stolterman, E. 2013. Design Research at CHI and Its Applicability to Design Practice. Proceedings of the SIGCHI Conference on Human Factors in Computing Systems (New York, NY, USA, 2013), 1951–1954.

[38] Roedl, D., Odom, W. and Blevis, E. 2017. Three Principles of Sustainable Interaction Design , Revisited. Digital Technology and Sustainability: Embracing the Paradox.

[39] Rogers, Y. 2004. New theoretical approaches for human-computer interaction. Annual Review of Information Science and Technology. 38, 1 (2004), 87–143.

[40] SDGs .:. Sustainable Development Knowledge Platform: 2017. https://sustainabledevelopment.un.org/sdgs. Accessed: 2017-03-10.

[41] Sharp, H., Rogers, Y. and Preece, J. 2007. Interaction Design: Beyond Human-Computer Interaction. Wiley.

[42] Shneiderman, B., Plaisant, C., Cohen, M. and Jacobs, S. 2009. Designing the User Interface: Strategies for Effective Human-Computer Interaction. Pearson.

[43] Silberman, M.S., Nathan, L., Knowles, B., Bendor, R., Clear, A., Håkansson, M., Dillahunt, T. and Mankoff, J. 2014. Next steps for sustainable HCI. interactions. 21, 5 (Sep. 2014), 66–69.

[44] Silberman, M.S. and Tomlinson, B. 2010. Toward an ecological sensibility: tools for evaluating sustainable HCI. CHI '10 Extended Abstracts on Human Factors in Computing Systems (New York, NY, USA, 2010), 3469–3474.

[45] Thomas, V., Remy, C., Hazas, M. and Bates, O. 2017. HCI and Environmental Public Policy: Opportunities for Engagement. Proceedings of the SIGCHI Conference on Human Factors in Computing Systems (2017).