Evaluation beyond Usability: Validating Sustainable HCI Research

1st Author Name
Affiliation
City, Country
e-mail address

2nd Author Name
Affiliation
City, Country
e-mail address

3rd Author Name
Affiliation
City, Country
e-mail address

ABSTRACT
The evaluation of research artefacts is an important step to validate research contributions. Sub-disciplines of HCI often pursue primary goals other than usability, such as Sustainable HCI (SHCI), HCI for development, or health and wellbeing. For such disciplines, established evaluation methods are not always appropriate or sufficient, and new conventions for identifying, discussing, and justifying suitable evaluation methods need to be established. In this paper, we revisit the purpose and goals of evaluation in HCI and SHCI, and elicit five key elements that can provide guidance to identifying evaluation methods for SHCI research. Our essay is meant as a starting point for discussing current and improving future evaluation practice in SHCI; we also believe it holds value for other subdisciplines in HCI that encounter similar challenges while evaluating their research.

Author Keywords
Sustainable HCI; Sustainability; Evaluation; Validation.

ACM Classification Keywords
H.5.2 Information interfaces and presentation (e.g., HCI): User interfaces—Evaluation/methodology

INTRODUCTION
In the HCI community, one important aspect for reviewers to consider is "the validity of the results you are presenting" [2]. This validity is often achieved through an evaluation process, such as an expert evaluation of a design idea or a user test with an interactive prototype [27,52,62,81,84]. Many of HCI's evaluation processes developed and matured over time; indeed, it took decades for HCI to agree on common evaluation standards, and the discussion about many of those standards continues to this day [26,28,34,82,83]. Alongside these discussions, new and emerging disciplines within HCI seek to adapt existing, as well as explore new, evaluation techniques for their research. However, existing evaluation processes have not always proven to be suitable for those new and emerging disciplines. Examples of disciplines struggling with evaluation processes are design fiction [53], information visualization [13], HCI for development (HCI4D) [5,11,21,50,51], and sustainable HCI [19,22,24,55,61,75,87]. In the remainder of this paper, we will focus on the latter.

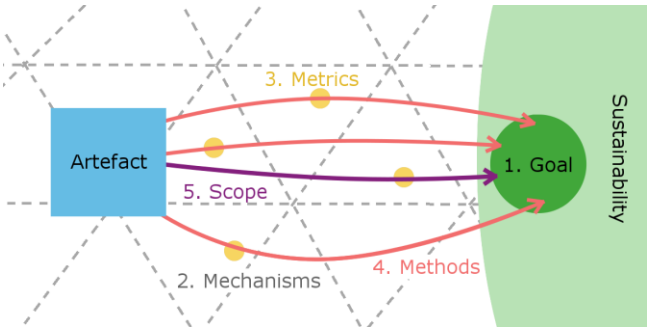


Figure 1. A model for eliciting the evaluation method for a research artefact in 5 steps. 1) Define the goal, 2) consider the surrounding mechanisms, 3) identify the metric for each mechanism, 4) find suitable method, 5) select scope.

Sustainable HCI (SHCI) is a relatively young field with its recent 10-year anniversary at CHI 2017 [14,57]. After an initial surge of research contributions, the growth of the field has recently slowed [9]. Some members of the SHCI community have been hoping to define its role and purpose to ensure its future [86]. We believe that the difficulties of evaluation – noted by various researchers [19,22,24,55,75,86] – present an obstacle for researchers looking to engage with the field. Being unsure of how to evaluate research can make it difficult for researchers to communicate the contributions or value of their projects.

Deciding how to evaluate an SHCI research contribution depends on the type of research conducted. Empirical studies that investigate practices of people can be sufficient for a contribution without further validation [28]. Similarly, frameworks and implications for design are acceptable research artefacts that traditionally are not further evaluated – also because such an evaluation would prove to be difficult [e.g., 26]. However, SHCI has come to a point where there is a variety of studies of practices, frameworks

for design, and implications for future research [24,77,79], but a rather low proportion of tangible solutions and hardly any evidence for a measurable impact on real-world practices [86]. Therefore, in this paper we focus on the evaluation of SHCI research artefacts–such as design interventions, interfaces, or even design ideas and sketches based on existing research–that aim to address issues related to environmental sustainability.

To explore the question of how to evaluate SHCI research, we first establish the background of evaluation in HCI, including its history, origins, and evolution. Revisiting past developments of evaluation processes and debates in SHCI gave us insights into what key elements are important for an evaluation process. Through an informal process of analysing those key elements and comparing them to scientific research about evaluation and validation outside of HCI we arrived at a recipe comprising five basic ingredients of evaluation: goals, mechanisms, metrics, methods, and scope. Reflecting on one's research with those five elements of evaluation in mind can provide guidance towards finding the appropriate evaluation method.

We discuss our model and its usefulness for SHCI research in solving problems of evaluation. Just as in traditional HCI, we believe there is no one-size-fits-all evaluation that can be applied to SHCI; our model is not a framework that, applied to any given research, immediately provides validation. Rather, we see these five ingredients as stepping stones for the community to engage in a debate about new avenues for evaluation. We envision the contribution of this paper to be a set of concepts for justifying and debating evaluation of HCI research *beyond usability*.

BACKGROUND

The topic of evaluation has a long-standing history in HCI and its related disciplines. In this section, we briefly recount this history and how its discussions have evolved. We also talk about the limitations and issues of evaluation in HCI that highlight how it is a moving target and in constant development as research advancements challenge existing views. In the second part, we summarize past efforts regarding evaluation in SHCI, including proposals to address the problem.

Evaluation in HCI

Human-Computer Interaction (HCI) has been defined as "the discipline concerned with the design, evaluation, and implementation of interactive computing systems for human use and with the study of major phenomena surrounding them" [40]. Despite advances in the field over the past 25 years, this definition is still up-to-date and in alignment with definitions in seminal HCI textbooks [27,52,81,84] and industry standards [44]. One major component in all those definitions is evaluation: "evaluation is integral to the design process" [81:584]; "we […] need to assess our designs and test our systems to ensure that they actually behave as we expect and meet the requirements of the user" [27:406]; "user-centred evaluation […] is a required activity in human-centred design" [44:16]. Thus, evaluation is a vital part of HCI; without evaluation, the validity of any research outcome may be threatened should other researchers struggle to determine if a research project has reached its desired goal or not.

Evaluation techniques have changed over time, as Barkhuus and Rode observed in their survey of 24 years of evaluation in CHI [7]. For example, there has been a shift towards more qualitative methods. Also, while the sample size of quantitative studies has decreased over time, the number of participants in qualitative studies has increased. The interdisciplinary nature of the field led to the adoption and use of new techniques, such as ethnographic studies, but also a stagnation of mixed methods studies. Moreover, formal evaluation of novel technological contributions is basically mandatory for CHI submissions; e.g., only 3% of surveyed papers had no evaluation in 2006.

Around the time that Barkhuus and Rode published their observations [7], a discussion emerged about whether technological innovations should be accepted without an evaluation of their usability aspects, or if they should at minimum offer a different kind of evaluation method. Buxton and Greenberg [34] argued that a prototype might perform poorly in an evaluation due to technological immaturity and different context compared to real-world practice, destroying a vision before it comes to full fruition. They argued that "the choice of evaluation methodology – if any – must arise from and be appropriate for the actual problem or research question under consideration" [34:119]. This statement applies to many new and emerging disciplines within HCI–including SHCI.

Since evaluation is such an integral part of HCI, the field has created a variety of different techniques to evaluate design artefacts, e.g., cognitive walkthrough, heuristic evaluation, think-aloud, task analysis, user questionnaires, just to name a few. Those techniques mostly cover how well the evaluated artefacts adhere to the standards defined in human-centred computing; in short, the techniques provide a usability evaluation. Due to its interdisciplinary nature, and the increased ubiquity of digital technologies, HCI has started to address more issues than just usability. Many HCI projects have endeavoured to address complex socio-economic issues, and those projects have demanded the fusion of previously disparate disciplines. This has created a diverse and exciting landscape of HCI research that is constantly evolving [54], just like HCI's evaluation methods [7]. But this diversity has been accompanied by evaluation challenges.

Evaluation challenges emerge not just because new disciplines find their way into HCI research, but also because different research environments and foci can bring different dimensions to research projects. For example, in HCI4D, there are technical, environmental, social, political, historical, and cultural dimensions that influence research

[11,21]. Moreover, the overriding goals in many HCI4D research projects are often not about usability but about aiding socio-economic development. These dimensions and goals have led researchers to argue that "HCI4D research must continue past evaluating the purely technical contributions" [5] and be more reflective of the unique circumstances in developing countries [50,51]. Although applying HCI's traditional usability evaluation methods can be worthwhile in some HCI4D contexts, the local environment and socio-economic development goals are equally if not more important to consider during evaluation.

Similarly, the goals and artefacts related to design fiction, as well as other future-focused and speculative research, can be difficult to validate using HCI's traditional usability evaluation methods [53]. Salovaara et al. [82] recently suggested that we need to rethink the way we evaluate future-focused prototypes in general. They argued for a new form of validation: beyond the "traditional criteria for empirical research – internal, construct, conclusion, and external validity" [82], prototypes should be subject to "projective validity". This projective evaluation requires an admittedly subjective justification of the prototype, including a definition of its intended futures. Such a projective evaluation might sound familiar to many SHCI researchers, as SHCI's ultimate goal is often future-focused change; the threat of climate change has made scientific communities well aware that our status quo cannot be maintained and we have to move towards a more sustainable future. Therefore, SHCI artefacts are often to be envisioned for use in the context of an uncertain or unknown future.

Evaluation in Sustainable HCI

Despite belonging to a relatively young field, the SHCI community realized the challenges of evaluation early on. In surveys of eco-feedback technology [32] and the field of SHCI in general [24], authors have recognized a lack of evaluations regarding sustainable aspects of technology interventions, noting that "many such papers are design descriptions" [24:1977]. This problem did not go unnoticed by the community, as two other contributions at the same conference presented early versions of a framework [22] and a toolbox [87] for evaluating SHCI. As researchers kept raising the issue of evaluation in subsequent years, the problem remained largely untouched: Brynjarsdottir et al. [19] state that of their 36 surveyed papers "almost half (17) have no user evaluation". One of the five patterns of Knowles et al. [48] targets the problem of evaluation, and a community effort to provide guidance towards next steps for the field lists SHCI research evaluation as one of the unanswered questions [86]. A preliminary taxonomy by Toyama [93] – refined by Lundström and Pargman [55] – marked the most recent attempt to solve this issue through a classification system for the sustainable impact of research.

One of the challenges related to evaluating SHCI contributions is the fact that the field can roughly be divided into two branches: sustainability in design (SiD) and sustainability through design (StD) [57]. Research under the umbrella of SiD is closest to Blevis's [14] seminal concept of Sustainable Interaction Design, targeting the design of products such that they are designed to have minimal impact. Work in StD aims to support sustainable lifestyles through the design of products that lead to a change in people's practices or awareness of sustainable issues. For many projects in StD an empirical evaluation would demand measuring behaviour change, but it is only a subset of work in this branch of SHCI. This distinction is particularly important for the discussion of evaluation in SHCI: StD artefacts usually resemble technology interventions that target the present to provoke change towards a more sustainable future; SiD targets the design of products which do not exist yet and whose manufacturing and use lies in a less proximate future.

The overwhelming majority of SiD contributions in SHCI represent studies about products from the past leading towards lessons learned for future design in the form of frameworks, design implications, and guidelines [77,79]. Such works are already self-sufficient as contributions and do not require further evaluation if the underlying research advances knowledge in the field and serves as inspiration for future research [28]. If we consider those design implications and guidelines and ask for an empirical evaluation, we encounter an unsurmountable challenge: "how can we know whether it is the theory (design principle, architecture, etc.) that was the cause, or whether it was the skill (or otherwise) of the designer" that yielded the effect measured in the evaluation [26]? As Dix states: "It is not just hard, but impossible" to tell. He suggests combining evaluation and justification (evidenced reasoning) to achieve validity, or in some cases eschewing any attempt at empirical evaluation, and instead validate the work solely through justification, as is done in mathematical proofs. Here we borrow the concept of projective validity by Salovaara et al. [82], envisioning the future our artefact is to be used in (which is still imaginary, similar to design fiction [17,88] or futures studies [58,65]) and justifying its validity based on historical evidence and past research.

JUSTIFICATION AS EVALUATION

Despite differences in how the evaluation will be carried out in detail for research in SiD or StD, there are commonalities that we can derive from past discussions of evaluation in SHCI. Silberman et al. [86] stress that research needs to define sustainability and develop the evaluation on a project-by-project basis; however, it can be paralyzing to put the entire burden of evaluation on the shoulders of researchers who already put their effort into attempts to solve issues of sustainability; especially as the field continues to reach out to new researchers [9]. A problem domain (e.g., climate change, HCI4D) may provide strong motivation for a piece of work, yet clearly, measuring a change in the large-scale multifaceted systemic

problem at the point and scale at which the research is conducted is infeasible.

This is not simply a question of burden to the researcher but rather a more fundamental question of where to draw the boundary around the evaluation, where to focus, how and what to measure, and where. In considering key elements of evaluation, we need to further elaborate on why evaluating such designs and systems at the point at which the work is conducted is so problematic. In short, we must address the following questions:

Q1. How do you validate an artefact in the present given the uncertainty of its future context of use?

In SiD or StD designs we intend to construct artefacts with repercussions external to the artefact itself. Necessarily, we even project forward toward an alternate future where the artefact and its adoption have 'come to pass'; many SHCI artefacts, like theories, are generative [26]: their real power lies in their application to real-world problems. In the case of SHCI, the effects of successful designs would only be potentially observable in a future where such a system has widescale adoption; where small-scale effects are multiplied and not obviated by other perhaps unforeseeable consequences or rebound effects [36,46]. This suggests that not only are the consequences of a design unknowable at the point at which the work is conducted, but that evaluation must be an ongoing and reflective process. In the interim, evaluation can only be done at a formative scale, and provide the initial justification for a design.

Q2. How do you justify a theory that unveils its real power through application?

As Dix points out regarding HCI theory, it is hard to validate generative artefacts, we can no longer rely on pure evaluation, but rather a nascent 'justification as evaluation' that provides insight toward the efficacy of the approach. We must construct a process that is a mixture of evaluation and justification [cf. 26], by building from and contributing to established theories and evidence, we can construct a strong chain of reasoning. While clearly, it would be rash to imply strong causality from a study to implied effect via such a chain of argument – the 'ripples of a design's adoption' becoming progressively more speculative and harder to measure – but we can contribute new theory and evidence on which to build. As with any evaluation in HCI, we must be both rigorous and cautious about both claiming and attributing presence or absence of effects. The evaluative scope of the justification may simply not include the participants or appropriate environment for which the design would be effective. We must be careful not to close off avenues of design too early.

Q3. How do you evaluate something that is embedded in a larger societal and ecopolitical system?

A particular challenge for SHCI centres around attribution or anticipation of causality for HCI. We have to ask, does a particular sustainable design 'work'? This is problematic for SiD and particularly StD, whose goals are to lead to intentional wider systemic change extrinsic to the artefact. Designs are framed as part of an ongoing and evolving discourse about how society, economy and governance address major global challenges. Such designs have their place in co-constructing a different future by challenging and reshaping how society anticipates and responds to these challenges. There is clearly a mismatch of scale and of presumed effect between these intended impacts and attribution to any single design; yet, iconic designs may well be pivotal in changing thinking. This is not an argument for laziness on behalf of the SHCI researcher, rather we must recognise such attribution of effect cannot be measured at the point of conception, if at all. We must focus instead on providing a credible, considered and ecologically valid justification for our work. We should expect the link to societal scale to remain rhetorical.

To make the process around evaluation in SHCI more concrete, our goal is then to help move this debate forward by evolving a set of evaluative ingredients that help us address this key challenge of how to identify the scope of what and where to evaluate. We also aim to offer metrics and methods for doing so, given the necessarily broad extrinsic goals an artefact may have. To provide more concrete guidance towards clearly identifying and justifying the evaluation for a given project, we suggest a recipe based on five key ingredients: goals, mechanisms, metrics, methods, and scope.

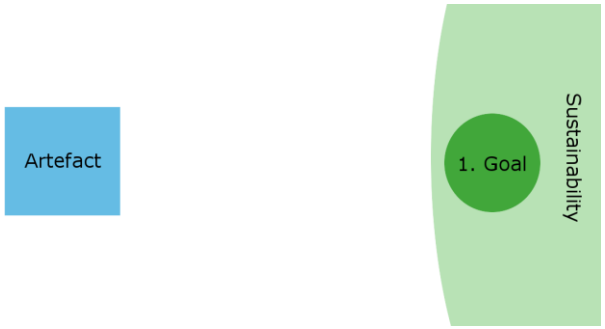
THE FIVE INGREDIENTS OF EVALUATION

Drawing on our assembled corpus of SHCI papers containing evaluation or discussion thereof (42 papers) and HCI literature (21 papers and 7 books) to date, we also broadened our consideration to wider evaluation literature from other fields (17 references) such as philosophy, sociology, or psychology. These 87 sources are referenced in the present paper; not explicitly mentioned are other sources we consulted in the process that contributed to our thought process, as well as conversations with members of the HCI community. The authors of this paper then applied an iterative process of reading, reflecting, and discussing our perspectives on the literature, and five elements solidified as key elements comprising the ingredients of our evaluation recipe.

For each of the five ingredients, a brief question highlights how it can be applied to a research project; when combined, all ingredients offer a recipe guiding towards an evaluation method. The introductory questions are accompanied by a description of how we envision their usage in a research process as well as general considerations. We contextualize each ingredient by elaborating on the takeaways for SHCI research based on the insights gained in the informal analysis process that led to this model.

1 – Goals

What is the goal that this SHCI research artefact is trying to achieve with regard to sustainability?



Specifically, this requires considering what aspects of sustainability are important within the scope of a project, and how the SHCI artefact addresses those aspects of sustainability. This goal is–or these goals are–supposed to be specific enough to elicit a general answer to a yes/no question about whether the project can be deemed successful following an evaluation; a generic "save the world" or "combat climate change" goal is not sufficient. As orientation, one might consider the Sustainable Development Goals (SDGs) [95] and formulate a concise, single sentence that connects the research artefact's impact to a concrete SDG-specific contribution. While the goal can potentially be iterated on in the subsequent design cycles that are typical in iterative HCI research [62], it should be high-level enough to stay the same such that the research can be iterated on with the same or similar goal in mind.

Project-specific goals instead of one general definition of sustainability

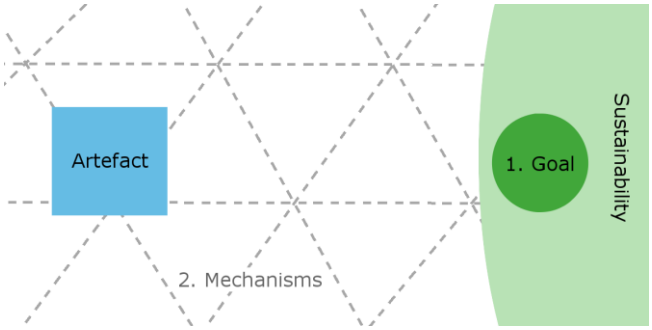
The SHCI community has attempted to create one definition of sustainability for the entire field [49,66,86]. However, differences in opinions about such a definition are exposed with simple questions such as whether sustainability is a process [86] or an endpoint [66]. Knowles and Håkansson [49] surveyed the community for a definition of sustainability and argued that a consensus on one single definition "is both unlikely and undesirable"; rather, definitions can vary based on researchers' personal interests, current work, and motivation. Therefore, the community concluded that "SHCI research should articulate clear study- or design-specific sustainability goals and metrics on a project-by-project basis" [86]. Defining such a goal not only provides a target at which researchers can aim to validate their work, it can also help to frame research and how it is assumed to contribute towards sustainability.

Usability evaluation: goals set the frame

Traditional evaluation in HCI offers clearly defined goals for assessing usability. Dix et al. [27:319] formulate "three main goals: to assess the extent and accessibility of the system's functionality, to assess users' experience of the interaction, and to identify any specific problems with the system". Preece et al. point to the fulfilment of users' needs [72:323], whereas Nielsen considers the overall quality and potential improvements of a system, as goals of evaluation [62:170]. These broad conceptions of goals are followed with more specific evaluation methods depending on what exactly is to be evaluated. In all of these cases, the goals are not meant to serve as templates for a simple question that can be answered instantly to finish an evaluation; rather, the goals set the frame for choosing which evaluation methodology to use. Nielsen [62] suggests that usability testing should start from writing a test plan and asking what is "the goal of the test: What do you want to achieve?" The same approach needs to be taken when evaluating SHCI research artefacts: usability (HCI) and sustainability (SHCI) is the overall target, but goals help researchers define what they want to achieve, and are necessary for developing a concrete evaluation plan.

2 – Mechanisms

What are the mechanisms that need to be considered to understand and assess the holistic impact of the research artefact in the context of external influencing factors?



It is important to consider how the artefact, applied to real-world practices and scenarios, will interfere with and be influenced by the "mechanisms" in its surrounding environment. Such mechanisms can be sociological or socio-technical in nature, which are important dimensions for StD research projects to consider if they aim to change the lifestyles and practices of people. Political and economic mechanisms on various scales play their role as well, especially for SiD research that addresses current and future economic issues, such as by visualizing supply chains [18] or asking HCI to consider labour conditions [30,60]. While political and economic mechanisms address a larger scale of impact, they are oftentimes out of the scope of influence for researchers.

Identifying the mechanisms that influence an SHCI artefact or project can be a difficult task. Indeed, researchers in the SHCI community have acknowledged this difficulty for quite some time [14,15,19,24,32] and have yet to completely resolve it. Calls to consider issues of scale [29] and abandon the pursuit of one-size-fits-all solutions [39,48,86] echo the difficulty of grappling with mechanisms. As such, some guidance might be helpful; such guidance can be found by revisiting the broader definitions provided by the SDGs [95] or the three pillars of sustainability [38:25]. For more inspiration or even concrete

suggestions for mechanisms, utilizing guidelines and frameworks from the large corpus of SHCI research (e.g., from surveys of StD [32,68] or SiD research [77,79]) or related other fields might be worthwhile.

Learning from other fields: the importance of mechanisms

The concept of "mechanisms" is used in various disciplines to describe the complex relationships between artefacts, individuals, societies, and nature. In philosophy, Popper argues that scientific theories can only be tested indirectly, requiring researchers to be aware of, understand, and observe the mechanisms at hand [69]. In behavioural science, structural equation modelling uses statistical methods to analyse relationships that contain unknown variables [47], thereby modelling mechanisms to gain a coherent picture of the relationships. And in fields such as biology, chemistry, and physics, experimental studies are historically validated by describing, observing, and analysing the mechanisms between an element and its environment (for direct validation) or changes in the environment only (for proof by implication).

In HCI, Dix [26] uses the term "mechanism" when discussing how to evaluate theory and theoretical frameworks. He refers to "mechanisms" as "the details of what goes on, whether in terms of user actions, perception, cognition, or social interactions". A common approach for evaluating theoretical frameworks in HCI is to have designers apply the frameworks to their practice and evaluate the outcome. This allows the designers to bypass any discussion of the mechanisms that might influence their design, as the resulting prototype can be assessed using traditional usability methodology; but such an approach is not always possible in SHCI because its theories or prototypes are often intended for a distant future and different context. Therefore, we argue SHCI needs to learn from other fields and go beyond its parent field in assessing the validity of research, by understanding a project's unique mechanisms and using them to justify its solutions [26].

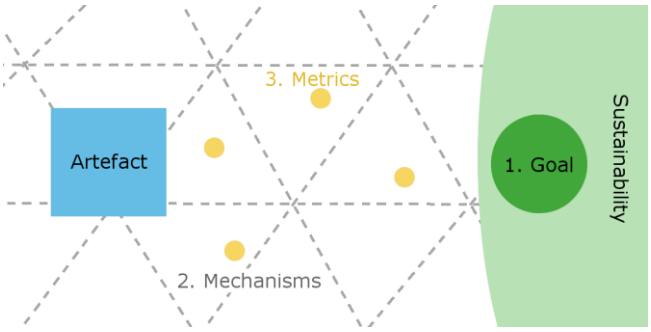
Acknowledging the big picture when justifying a design

Broader societal mechanisms, such as economic [16,23,30,60,66] and political [29,30,59,60,91,92] should be considered by SHCI research as well, but are more difficult to evaluate. An empirical evaluation is usually not possible since a single design idea or prototype will not result in any measurable change of the large-scale political economy. We argue that those mechanisms are still important to identify for two reasons. First, clearly formulating large-scale political, economic, and socio-technical implications might help to elicit related small-scale mechanisms that are suitable implications for an evaluation. Second, when evaluating an artefact by justification [26,82] acknowledging those mechanisms in projected future scenarios shows a deep level of reflection and awareness.

An example for this is the rebound effect [36,46], an economic observation stating that an increased efficiency of technology is met with an increased rate of consumption due to higher demand. Rebound effects are long-term implications of large-scale technological developments and therefore unrealistic to measure in an empirical evaluation of SHCI research; yet, acknowledging them is a better strategy than waiting for reviewers to bring up those arguments and potentially question whether *not* designing technology would have been the better choice [10,67].

3 – Metrics

What are the metrics that can be observed or measured to assess any given mechanism with regard to reaching the desired goal?



For each mechanism that a researcher identifies and considers for evaluation, the corresponding metrics have to be defined. Metrics can take on different forms, such as a quantifiable amount of resources, practices of people affected by the artefact, or the opinion of experts. Sometimes metrics may be similar to the goal. However, a goal usually defines an endpoint and the evaluation addresses whether or not the goal was reached, whereas metrics provide a spectrum to elaborate on how well the goal was reached – or by how much it was missed.

Choosing among multiple metrics for a mechanism

Even when a goal is clearly defined and the mechanisms have been identified, SHCI research might fail to validate properly if the most suitable metrics are not selected. Oftentimes researchers want to measure long-term impact, such as the satisfaction [98] or attachment [33,63] between consumers and devices in the case of SiD, or behaviour change for StD [19]. Directly assessing those effects is not always feasible or in some cases even impossible (such as attachment that develops over decades), which is why alternative metrics need to be found that can help justify how the artefact influenced the respective mechanism. For measuring behaviour change, Knowles et al. [48] recommend to instead look at "impact ripples", i.e., comparing other factors such as participants' attitudes before and after the technology intervention.

One of the more concrete methods for assessing the impact on environmental sustainability is Life Cycle Assessment (LCA) [31,41,43] which provides a range of metrics regarding the entire lifecycle of a product, with several input and output variables depending on the underlying LCA database and software. Some work in SHCI has used

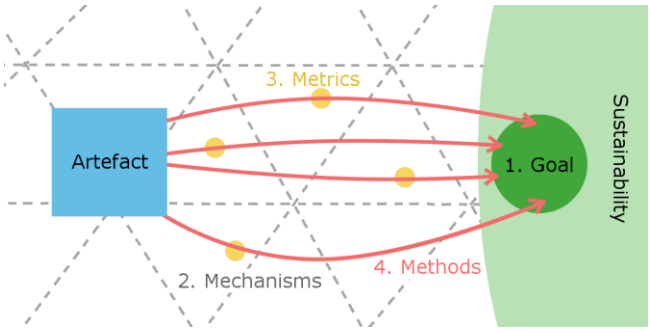
LCA data [8,12,18]. A comprehensive LCA database could be as close as it gets as a reliable repository for many different SHCI metrics; especially for direct resource impact projection of SiD. Another international standard that can be useful in deriving those is the ISO for Environmental Technology Verification [45] that offers metrics to assess the performance of environmental technologies. The few existing evaluation frameworks in SHCI to assess sustainability offer criteria [22], three different levels [93], and various dimensions [55] that can either serve as concrete metrics for projects, especially in SiD [22], or at least provide guidance towards identifying potential metrics for an evaluation.

Combined metrics as the ultimate goal?

Despite the goals centring around sustainability, usability should not be neglected entirely when designing technology interventions, therefore rendering traditional usability metrics still useful. The most desirable evaluation for an SHCI artefact combines sustainability and usability aspects. As such, adapting and extending HCI metrics by sustainable criteria might be a goal worth pursuing. For example, SHCI could extend usability heuristics or usability principles [27,52,62,81,84] by adding sustainable goals into them. Similarly, Froehlich et al. [32] contrast how environmental psychology offers guidelines to measure behaviour change while HCI provides means to evaluate traditional usability criteria. Ultimately, combining those would not only be one step towards addressing the evaluation issues in the field of SHCI, but also fulfil Blevis's vision of making sustainability "a central focus of interaction design" [14].

4 – Methods

What methods can be employed to assess the metrics for the respective mechanisms to answer the question to what extent the artefact has reached the desired goal?



Identifying mechanisms is critical for understanding the relationships between an artefact and its complex environment, and identifying metrics provides the interface for researchers to investigate such mechanisms. For conducting the evaluation, however, a researcher needs to decide how to observe or measure such metrics. This is where identifying the assessment method comes in. Assessment methods can either be existing or new evaluation techniques, chosen or adapted from the rich toolbox of methods in SHCI, HCI, and other fields. They can take on many different forms: quantitative or qualitative, lab or field, empirical or theoretical, long-term or short-term, large-scale or small sample, with participants or designers, and so on. This step is not about conducting the evaluation itself; it is about surveying the available options and selecting the most appropriate one for any given mechanism and metric.

Need for methods that deal with uncertainty and future implications

For reviewing the choice of methods in SHCI, we need to clearly distinguish between SiD and StD. In SiD, there is a glaring lack of methods for evaluating SiD artefacts, but also a lack of SiD artefacts themselves as most contributions are of theoretical nature [77,79]. Evaluating such theories is an almost impossible task [26], but unless they are applied to practice an evaluation is also not required for submitting such research; therefore, we argue that the lack of evaluation might be one of the obstacles to move the field forward. Although Blevis explicitly discusses a possibility for evaluating SiD through design critique [15] and using his rubric [14], we can only find two contributions at CHI that apply SiD to practice and attempt an evaluation [37,76].

For StD, the field has seen more variety and application of evaluation methods, although several critiques argue that the efforts so far were insufficient in assessing the true impact of the created artefacts [19,29,32,48]. While many of those critiques address the stages of identifying appropriate mechanisms and metrics, broadening the vision of available methods can be helpful here as well. Brynjarsdottir et al. [19] report that one third (12 out of 36) of their surveyed papers cannot evaluate because "they explore a design methodology or are in the early stages of design". Their recommendation of using participatory design would address such issues as it offers evaluation in the design process and not just at its completion.

Despite those different obstacles for identifying appropriate evaluation methods, there is one commonality: assessing the validity of sustainability requires an assessment of uncertain future implications. But SHCI cannot shy away from evaluation because of the immature state of prototypes and research artefacts. The field has to find methods that can evaluate prototypes early on (e.g., for persuasive technology), or even when the solution is only a sketched design idea (e.g., for sustainable design of products). Such methods need to estimate an impact, draw scenarios of future trajectories [73], or discuss the uncertainties [56]. Research contributions that discuss the development of evaluation in HCI can help SHCI to identify new methods, such as by focus on problem-solving [64], using projective validity [82], reconsidering what validity is "about" [25], or focusing more on the justification of theories [26].

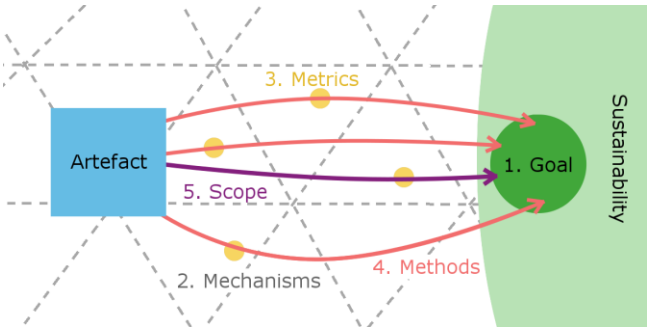
Looking beyond HCI for tools and methods

Thanks to its interdisciplinary nature, HCI has historically adapted evaluation techniques from other fields and used them to investigate usability aspects, and SHCI needs to do the same. For StD, where the focus of evaluation is more concerned about the sociological implications of technology, a deeper look into psychology (as recommended by Froehlich et al. [32]) or social practice theory [85] can be most promising. Design science research, a field closely related to HCI in its goals and methodology, offers a rich body of research about evaluation methods by classifying frameworks [96,97], strategies [74], or evaluation criteria [70], serving for further inspiration to arrive at appropriate evaluation methods.

The previously discussed industry standards for LCA [43] and Environmental Technology Verification [45] also are most useful in eliciting metrics but offer guidance towards methods as well. For example, while the LCA database offers metrics to measure the environmental impact of products, LCA software such as SimaPro [71] or OpenLCA [35] are the associated tools that calculate the result and answer the questions asked in an evaluation process. Those resources might be most useful for SiD in which the technology or theories informing technology are to be evaluated.

5 – Scope

What is the scope of mechanisms that can, should, and will be considered for validating the artefact?



It is normally impossible to pay justice to every mechanism involved, therefore it is critically important to set the boundaries of which mechanisms are chosen for the evaluation process. The selection process should be transparent and justified by well-constructed arguments, and is predominantly guided by two factors: relevance (which mechanisms are most suitable for assessing whether the research reaches its desired goal) and feasibility (which mechanisms can be evaluated given the constraints of time and resources typically associated to a research project).

While narrowing one's focus and excluding mechanisms can be done right after identifying them, it might be advisable to go through the options for metrics and methods for each mechanism. If a metric cannot be defined or is deemed to be impossible to be assessed because there is no feasible method, the scope can be adjusted and the associated mechanism will not be a candidate for evaluation. In such a case, the absence of the metric or method serves as justification for why the mechanism was excluded, and the believability of the evaluation depends on how well the options for potential metrics and methods were explored.

When to broaden out and when to narrow down in SHCI

While researchers in SHCI extensively discuss the importance of acknowledging the complexity of external mechanisms as outlined earlier, there is not much explicit advice to narrow such scope. We believe this to be a symptom of a relatively young field that is eager to grow and therefore asks for expanding its horizons [9,42,86], but also a symptom of the lack of established evaluation methods. However, there is an important difference between identifying a project, its goal, and suitable solutions (at which stage exploration and expansion is the best move) and evaluating the conducted research with scientific rigor (which requires a good focus).

We propose a re-interpretation of common advice given in some of the most prominent SHCI critiques [19,24,48,86] who urge SHCI research consider the mechanisms and the complexity of societal systems. This can be understood as adding more dimensions and therefore further complicating the evaluation; however, we suggest replacing, or re-focusing, the evaluation process. For example, for StD we take the advice from Brynjarsdottir et al. [19] to "shift from behaviours to practices"; instead of measuring behaviour to validate artefacts, researchers might analyse related practices. While traditionally such a practice-oriented approach suggests broadening the perspective [85], in SHCI it can be used to shifting one's scope from evaluating the individual in the moment of interaction – a practice HCI researchers are familiar with from usability evaluations – to more sociological approaches of taking the practice as unit of analysis and considering broader social and institutional arrangements.

Focus on one evaluation at a time

The closest analogy for highlighting the importance of choosing the right scope can be found by looking at quantitative evaluation methodology in a field like psychology, where there are many unknown variables and therefore reducing the scope is mandatory to prove statistical significance. Other sciences, such as complex systems theory, draw boundaries around networks and organizations [6,20] to even consider the complex systems observable as such. Alexander, whose pattern language intended to describe the complexity of architecture from small-scale rooms in a building up to large-scale patterns of towns [4] reminds the reader to apply "one pattern at a time" [3]: "When we have the order of the language right, we can pay attention to one pattern at a time, with full intensity". The same holds true for SHCI evaluation: we need to be aware of the entire complex system that is at

stake, but when it comes to applying our artefact to real-world practice and validating its impact, it is imperative to focus on one mechanism at a time. And the better our understanding of the system, the better we can separate its mechanisms, and the more confident we can be in analysing and evaluating it.

DISCUSSION

In this section, we discuss the broader implications of our five ingredients for the debate about how to evaluate research artefacts beyond usability. Specifically, we elaborate on how our evaluation model can enrich the debate and move the field forward. We also reflect on the theory-practice gap, and how our recipe for evaluation might be used to address problems of evaluation in other disciplines within HCI.

Opportunities and Challenges of our Recipe for Evaluation

Thanks to its open-endedness and generativity, applying our recipe to SHCI research should result in several ways of evaluating a project or artefact; any combination of mechanisms, metrics, and methods could lead to a unique evaluation. As such, our recipe offers several opportunities: researchers can select an evaluation method they are familiar with, confident in, and have the resources available for. Being aware of the alternatives and setting them in relation to the mechanisms also facilitates the justification of why a particular evaluation method was chosen. Furthermore, the recipe can be applied in the early stages of research to plan ahead, taking necessary precautions or accommodating for additional data-tracking by measuring relevant metrics from the start.

However, we also anticipate challenges in applying this recipe as it is not an evaluation itself, and therefore its success depends on the researchers' knowledge and expertise of relevant mechanisms, metrics, and methods. For example, identifying mechanisms is probably the most difficult task as there is no clear guidance for how to find mechanisms. One can start by branching out from general dimensions of sustainability such as the SDGs [95] or the three pillars [38], but in the future one contribution of the SHCI community could be to establish a taxonomy or classification system of mechanisms. Just as one can pick from a multitude of usability evaluation techniques by surveying HCI textbooks, we envision a similar collection of mechanisms as well as related metrics and methods for SHCI. Silberman and Tomlinson [87] recommend developing "principles, heuristics, and indices" as next steps for the community to evaluate SHCI research. We echo their call and suggest focusing on identifying mechanisms and metrics first, as those will guide towards methods for evaluation.

Another point we want to emphasize is that we present only "a" model for evaluation, but it is neither the final nor the only one. Other abstractions that formalize the evaluation process might have more or different elements, and they might lead to similar results. Our model does not end the problem of evaluation in SHCI – but it is a first step towards it and will hopefully provoke the community to expand on the idea, leading to a rich corpus of evaluation methodology. There is much room for improvement, and especially in the SiD branch we see evaluation largely unexplored; this is likely caused by the fact that much of SiD works on an emotional level [63,98] which makes any evaluation difficult. As Dix writes, validating research "is even more problematic when the systems we design are intended to elicit emotions, to be fun, to yield experiences. These things take their validity from their subjectivity" [25]. SHCI has to take on that challenge and acknowledge that new evaluation methods will rely on subjectivity, or as Sengers and Gaver describe: "Evaluation is also a form of interpretation" [83:105].

Rethinking Evaluation and what it is about in SHCI

We consider our discussion of the problem of evaluation in SHCI in this paper only a starting point for the community to create more concrete solutions and move forward. The five elements we identified contribute to the debate by making it more tangible and accessible. We intentionally decided to not provide concrete examples for the five elements; the contribution of this paper is not supposed to be a model of evaluation, but a rethinking of how we evaluate SHCI research. This comes at the risk of exposing our recipe and its ingredients to many different interpretations – however, we perceive this not as a risk but as an opportunity of redefining, rethinking, and remodelling evaluation. As was pointed out in previous research about validating theories: the best way to evaluate a theory is not its application to an example (which leads to an evaluation of the example or its creator rather than the theory) but its justification [26].

Nevertheless, we call for the community to attempt to apply the ingredients to their research, be it in theoretical discussions that contribute to this debate or in concrete examples in the form of case studies with lists of goals, mechanisms, metrics, methods, and scope. Such case studies and lists of example evaluations could become the basis of a community effort to address issues of evaluations, helping researchers to choose from a knowledgebase of approved solutions. It would also benefit reviewers and editors of workshops, conferences, and journals to justify and recognize sustainability work. This could help reduce the frequency with which SHCI research falls through the cracks for not meeting the standards set by traditional HCI evaluation methods.

The threat of sustainability work not being able to adhere to conventional standards of HCI research due to the unresolved evaluation problem is particularly daunting. Since the world's leaders recognized "the need for an effective and progressive response to the urgent threat of climate change on the basis of the best available scientific knowledge" [94], HCI should not exempt itself from

contributing to this scientific knowledge. The internationally recognised urgency of addressing climate change speaks to the gravity of the issue: SHCI cannot wait decades to come to an agreement about how to evaluate its research contributions – climate change is an issue that needs to be addressed right here, right now. But climate change is too broad of a concept to be addressed directly; hence the *goals* and *metrics* ingredients in our recipe urge SHCI researchers to think in more tractable terms.

At the same time, it is important to consider the big picture and keep in mind what research in SHCI is about, and *mechanisms* help to make that connection. Drawing upon Abrams' theory of a universe with different focuses [1], Dix [25] notes that research is often *about* multiple things. An eco-feedback display may be *about* reducing the energy footprint of a household, but likewise it is *about* achieving a long-term impact on people's practices as well as *about* combatting climate change. The discussion of scale in SHCI is often interpreted as one that distinguishes "good" and "bad" approaches to those problems, but we suggest reminding ourselves as an SHCI community that all those mechanisms together form a system (that scales almost indefinitely). As researchers, we need to adjust our *scope* appropriately to be able to evaluate our work using *methods* that are feasible. If we as SHCI researchers can validate our contribution that is *about* impact on a small scale, we can build outwards based on the understanding of the system's complexity to justify how this contribution factors into the intractable problem that is *about* climate change.

Connecting Research and Practice: Evaluating Frameworks

While we focus our efforts solely on the field of SHCI in this paper, we believe there are lessons for other disciplines in HCI and even the field in general. Despite the rich history and availability of evaluation methodologies, some argue that evaluation "has not been systematically studied" [7] and several contributions demonstrate the need for a more nuanced debate. Arguments about whether to evaluate or not [34] might be a symptom of even the well-established field of HCI not having the necessary tools for evaluating all its research, e.g., when its output is more theoretical [26,82].

As a result, research often ends with design implications, frameworks, and guidelines; useful research contributions that inform the practice of future technology without doubt. However, many of those frameworks are almost never put to practice, which has been described as the theory-practice gap [78,80,89,90]. In SiD specifically, there exist several frameworks and guidelines as recent surveys show [77,79] but "sustainable HCI research has had little impact outside HCI" [86]. If SHCI wants to contribute to scientific knowledge and also achieve real-world impact, it is time for the field to switch from producing more frameworks and design implications to putting them into practice. We see two opportunities arise: first, as Silberman and Tomlinson [87] allude to, those frameworks might inform the heuristics and principles that in the future lead to new evaluation methods. Second, if the frameworks are not put into practice directly, perhaps we could think of ways to evaluate the frameworks and come to an agreement which to focus on in moving SHCI forward.

Potential for other Application Areas in HCI

As mentioned earlier, other fields within HCI have reported issues related to evaluation. Our model might serve as inspiration for a solution in those domains, too. For example, one field that shares similarities with SHCI is design fiction, whose artefacts are occasionally similar to speculative SiD design concepts. Since the evaluation process for design fiction artefacts has to be conducted with a constructed future in mind, design fiction researchers rely heavily on identifying relevant mechanisms and justifying the design rather than an empirical validation. Therefore, our model could be seen as the glue to connect the issues of design fiction [53] and concepts of projective validity [82], evaluation by justification [26], and interpretation [83].

Another field that shares some similar goals and concerns with SHCI is HCI4D, which has been aware of its evaluation problems for longer than SHCI has even existed [5,11,21,50]. That this problem persists despite years of awareness suggests that a different approach to evaluation might be necessary. To grapple with the country-specific, complex technical, environmental, social, political, historical, and cultural dimensions that influence HCI4D research [11,21] means to acknowledge the mechanisms at play, and therefore our model might be of value here, too.

CONCLUSION

In this paper, we sought to explore ways to address the evaluation problem when the primary goal is not usability. We have looked at the history of evaluation in HCI and SHCI and discussed ways to assess the validity of sustainability in research projects. As a starting point for and to facilitate the debate within the community we highlighted five ingredients that, based on past research on evaluation in HCI and other fields, are critical for identifying an appropriate evaluation method. We emphasize that our recipe is not meant as a framework for evaluation itself, but as a process to guide researchers towards identifying the most suitable evaluation for their research; and to give researchers a common set of terms for justifying and debating evaluation. In our understanding, this work is only a first step towards solving the evaluation problem, and while we believe that it is a helpful step not only for SHCI but also for other areas in HCI, more work is required. We hope that the community builds on this work and that it spurs the debate about identifying new ways to evaluate research artefacts beyond usability.

REFERENCES

1. Meyer H. Abrams. 1953. *The Mirror and the Lamp: Romantic Theory and the Critical Tradition*. Oxford University Press, New York.
2. ACM SIGCHI. 2017. Guide to a Successful Paper or Note Submission. Retrieved from https://chi2017.acm.org/guide-paper-submission.html
3. Christopher Alexander. 1979. *The Timeless Way of Building*. Oxford University Press, New York, NY.
4. Christopher Alexander, Sara Ishikawa, and Murray Silverstein. 1977. *A Pattern Language: Towns, Buildings, Construction*. Oxford University Press, New York.
5. Yaw Anokwa, Thomas N. Smyth, Divya Ramachandran, Jahanzeb Sherwani, Yael Schwartzman, Rowena Luk, Melissa Ho, Neema Moraveji, and Brian DeRenzi. 2009. Stories from the Field: Reflections on HCI4D Experiences. *Information Technologies & International Development* 5, 4: 101–116.
6. Holly Arrow, Joseph Edward McGrath, and Jennifer L. Berdahl. 2000. *Small Groups as Complex Systems: Formation, Coordination, Development, and Adaptation*. SAGE Publications, Inc, Thousand Oaks, Calif.
7. Louise Barkhuus and Jennifer A. Rode. 2007. From Mice to Men - 24 Years of Evaluation in CHI. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems* (CHI '07). https://doi.org/10.1145/1240624.2180963
8. Oliver Bates, Mike Hazas, Adrian Friday, Janine Morley, and Adrian K. Clear. 2014. Towards an Holistic View of the Energy and Environmental Impacts of Domestic Media and IT. In *Proceedings of the 32Nd Annual ACM Conference on Human Factors in Computing Systems* (CHI '14), 1173–1182. https://doi.org/10.1145/2556288.2556968
9. Oliver Bates, Vanessa Thomas, and Christian Remy. 2017. Doing Good in HCI: Can We Broaden Our Agenda? *Interactions (New York)* 24, 5.
10. Eric P.S. Baumer and M. Six Silberman. 2011. When the Implication is Not to Design (Technology). In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems* (CHI '11), 2271–2274. https://doi.org/10.1145/1978942.1979275
11. Jack Beaton and Ripul Kumar. 2010. Indian Cultural Effects on User Research Methodologies. In *CHI '10 Extended Abstracts on Human Factors in Computing Systems* (CHI EA '10), 4267–4272. https://doi.org/10.1145/1753846.1754137
12. Peter Beigl, Stewart Hickey, Gudrun Obersteiner, Colin Fitzpatrick, Karsten Schischke, Paul Maher, and Jose Ospina. 2013. Towards zero waste in industrial networks: a case study of the D4R laptop. In *First International Conference on ICT for Sustainability* (ICT4S '13).
13. Enrico Bertini, Catherine Plaisant, and Giuseppe Santucci. 2007. BELIV'06: Beyond Time and Errors; Novel Evaluation Methods for Information Visualization. *interactions* 14, 3: 59–60. https://doi.org/10.1145/1242421.1242460
14. Eli Blevis. 2007. Sustainable interaction design: invention & disposal, renewal & reuse. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems* (CHI '07), 503–512. https://doi.org/10.1145/1240624.1240705
15. Eli Blevis, Youn-kyung Lim, David Roedl, and Erik Stolterman. 2007. Using design critique as research to link sustainability and interactive technologies. In *Proceedings of the 2nd international conference on Online communities and social computing* (OCSC '07), 22–31. Retrieved March 25, 2013 from http://dl.acm.org/citation.cfm?id=1784297.1784301
16. Eli Blevis, Kevin Makice, William Odom, David Roedl, Christian Beck, Shunying Blevis, and Arvind Ashok. 2007. Luxury & new luxury, quality & equality. In *Proceedings of the 2007 conference on Designing pleasurable products and interfaces* (DPPI '07), 296–311. https://doi.org/10.1145/1314161.1314188
17. Mark Blythe. 2014. Research Through Design Fiction: Narrative in Real and Imaginary Abstracts. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems* (CHI '14), 703–712. https://doi.org/10.1145/2556288.2557098
18. Leonardo Bonanni, Matthew Hockenberry, David Zwarg, Chris Csikszentmihalyi, and Hiroshi Ishii. 2010. Small business applications of sourcemap: a web tool for sustainable design and supply chain transparency. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems* (CHI '10), 937–946. https://doi.org/10.1145/1753326.1753465
19. Hronn Brynjarsdottir, Maria Håkansson, James Pierce, Eric Baumer, Carl DiSalvo, and Phoebe Sengers. 2012. Sustainably unpersuaded: how persuasion narrows our vision of sustainability. In *Proceedings of the 2012 ACM annual conference on Human Factors in Computing Systems* (CHI '12), 947–956. https://doi.org/10.1145/2208516.2208539
20. Paul Cilliers. 2001. Boundaries, hierarchies and networks in complex systems. *International Journal of Innovation Management* 05, 02: 135–147. https://doi.org/10.1142/S1363919601000312
21. Michael Demmer, Eric Brewer, R.J. Honicky, Madelaine Plauch, Melissa Ho, Sonesh Surana, and Joyojeet Pal. 2006. The Challenges of Technology Research for Developing Regions. *IEEE Pervasive Computing* 5: 15–23. https://doi.org/doi.ieeecomputersociety.org/10.1109/MPRV.2006.40
22. Tawanna Dillahunt, Jennifer Mankoff, and Jodi Forlizzi. 2010. A proposed framework for assessing

environmental sustainability in the HCI community. In *Examining Appropriation, Re-Use, and Maintenance of Sustainability workshop at CHI 2010* (CHI '10 workshop).

23. Tawanna Dillahunt, Jennifer Mankoff, Eric Paulos, and Susan Fussell. 2009. It's not all about "Green": energy use in low-income communities. In *Proceedings of the 11th international conference on Ubiquitous computing* (Ubicomp '09), 255–264. https://doi.org/10.1145/1620545.1620583
24. Carl DiSalvo, Phoebe Sengers, and Hrönn Brynjarsdóttir. 2010. Mapping the landscape of sustainable HCI. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems* (CHI '10), 1975–1984. https://doi.org/10.1145/1753326.1753625
25. Alan Dix. 2004. validity. In *CHI2004 Workshop on Reflexive HCI*.
26. Alan Dix. 2008. Theoretical analysis and theory creation. In *Research Methods for Human-Computer Interaction*. Cambridge University Press.
27. Alan J. Dix, Janet E. Finlay, Gregory D. Abowd, Russell Beale, and Janet E. Finley. 1998. *Human-Computer Interaction*. Prentice Hall, London ; New York.
28. Paul Dourish. 2006. Implications for Design. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems* (CHI '06), 541–550. https://doi.org/10.1145/1124772.1124855
29. Paul Dourish. 2010. HCI and environmental sustainability: the politics of design and the design of politics. In *Proceedings of the 8th ACM Conference on Designing Interactive Systems* (DIS '10), 1–10. https://doi.org/10.1145/1858171.1858173
30. Hamid Ekbia and Bonnie Nardi. 2015. The Political Economy of Computing: The Elephant in the HCI Room. *interactions* 22, 6: 46–49. https://doi.org/10.1145/2832117
31. European Commission: COM. 2015. Closing the loop - An EU action plan for the Circular Economy. Retrieved October 27, 2016 from http://eur-lex.europa.eu/legal-content/EN/TXT/?uri=CELEX:52015DC0614
32. Jon Froehlich, Leah Findlater, and James Landay. 2010. The design of eco-feedback technology. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems* (CHI '10), 1999–2008. https://doi.org/10.1145/1753326.1753629
33. Silke Gegenbauer and Elaine M. Huang. 2012. Inspiring the design of longer-lived electronics through an understanding of personal attachment. In *Proceedings of the Designing Interactive Systems Conference* (DIS '12), 635–644. https://doi.org/10.1145/2317956.2318052
34. Saul Greenberg and Bill Buxton. 2008. Usability Evaluation Considered Harmful (Some of the Time). In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems* (CHI '08), 111–120. https://doi.org/10.1145/1357054.1357074
35. Greendelta. 2017. openLCA.org. Retrieved September 10, 2017 from /
36. Lorna A. Greening, David L. Greene, and Carmen Difiglio. 2000. Energy efficiency and consumption — the rebound effect — a survey. *Energy Policy* 28, 6: 389–401. https://doi.org/10.1016/S0301-4215(00)00021-5
37. Barbara Grosse-Hering, Jon Mason, Dzmitry Aliakseyeu, Conny Bakker, and Pieter Desmet. 2013. Slow Design for Meaningful Interactions. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems* (CHI '13), 3431–3440. https://doi.org/10.1145/2470654.2466472
38. Jonathan M. Harris. 2000. *Basic principles of sustainable development*. Tufts University.
39. Helen Ai He, Saul Greenberg, and Elaine M. Huang. 2010. One size does not fit all: applying the transtheoretical model to energy feedback technology design. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems* (CHI '10), 927–936. https://doi.org/10.1145/1753326.1753464
40. Thomas T. Hewett, Ronald Baecker, Stuart Card, Tom Carey, Jean Gasen, Marilyn Mantei, Gary Perlman, Gary Strong, and William Verplank. 1992. *ACM SIGCHI Curricula for Human-Computer Interaction*. ACM, New York, NY, USA.
41. Lorenz M. Hilty. 2008. *Information Technology and Sustainability*. Books On Demand, Norderstedt Germany.
42. Elaine M. Huang. 2011. Building outwards from sustainable HCI. *interactions* 18, 3: 14–17. https://doi.org/10.1145/1962438.1962444
43. International Organization for Standardization. 2006. ISO 14040:2006 - Environmental management -- Life cycle assessment -- Principles and framework. Retrieved September 9, 2017 from https://www.iso.org/standard/37456.html
44. International Organization for Standardization. 2010. ISO 9241-210:2010 - Ergonomics of human-system interaction -- Part 210: Human-centred design for interactive systems. Retrieved August 25, 2017 from https://www.iso.org/standard/52075.html
45. International Organization for Standardization. 2016. ISO 14034:2016 - Environmental management -- Environmental technology verification (ETV). Retrieved August 25, 2017 from https://www.iso.org/standard/43256.html
46. Samuel J. Kaufman and M. Six Silberman. 2011. Rebound effects in sustainable HCI. In *Sustainable Interaction Design in Professional Domains workshop at CHI 2011* (CHI '11 workshop).
47. Rex B. Kline. 2010. *Principles and Practice of Structural Equation Modeling, Third Edition*. The Guilford Press, New York.

48. Bran Knowles, Lynne Blair, Stuart Walker, Paul Coulton, Lisa Thomas, and Louise Mullagh. 2014. Patterns of Persuasion for Sustainability. In *Proceedings of the 2014 Conference on Designing Interactive Systems* (DIS '14), 1035–1044. https://doi.org/10.1145/2598510.2598536
49. Bran Knowles and Maria Håkansson. 2016. A Sustainable HCI Knowledge Base in Progress. *interactions* 23, 3: 74–76. https://doi.org/10.1145/2904896
50. Neha Kumar and Susan Dray. 2017. Crossing Borders, Building Bridges. *interactions* 24, 3: 42–47. https://doi.org/10.1145/3058494
51. Neha Kumar, Naveena Karusala, Aaditeshwar Seth, and Biswajit Patra. 2017. Usability, Tested? *interactions* 24, 4: 74–77. https://doi.org/10.1145/3098571
52. Jonathan Lazar, Jinjuan Heidi Feng, and Harry Hochheiser. 2010. *Research Methods in Human-Computer Interaction*. Wiley, Chichester, West Sussex, U.K.
53. Joseph Lindley, Paul Coulton, and Emmett L. Brown. 2016. Peer Review and Design Fiction: "Great Scott! The Quotes Are Redacted." In *Proceedings of the 2016 CHI Conference Extended Abstracts on Human Factors in Computing Systems* (CHI EA '16), 583–595. https://doi.org/10.1145/2851581.2892568
54. Yong Liu, Jorge Goncalves, Denzil Ferreira, Bei Xiao, Simo Hosio, and Vassilis Kostakos. 2014. CHI 1994-2013: Mapping Two Decades of Intellectual Progress Through Co-word Analysis. In *Proceedings of the 32Nd Annual ACM Conference on Human Factors in Computing Systems* (CHI '14), 3553–3562. https://doi.org/10.1145/2556288.2556969
55. Anton Lundström and Daniel Pargman. 2017. Developing a Framework for Evaluating the Sustainability of Computing Projects. In *Proceedings of the 2017 Workshop on Computing Within Limits* (LIMITS '17), 111–117. https://doi.org/10.1145/3080556.3080562
56. Jennifer Mankoff. 2012. HCI and sustainability: a tale of two motivations. *interactions* 19, 3: 16–19. https://doi.org/10.1145/2168931.2168937
57. Jennifer C. Mankoff, Eli Blevis, Alan Borning, Batya Friedman, Susan R. Fussell, Jay Hasbrouck, Allison Woodruff, and Phoebe Sengers. 2007. Environmental sustainability and interaction. In *CHI '07 Extended Abstracts on Human Factors in Computing Systems* (CHI EA '07), 2121–2124. https://doi.org/10.1145/1240866.1240963
58. Jennifer Mankoff, Jennifer A. Rode, and Haakon Faste. 2013. Looking Past Yesterday's Tomorrow: Using Futures Studies Methods to Extend the Research Horizon. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems* (CHI '13), 1629–1638. https://doi.org/10.1145/2470654.2466216
59. Samantha McDonald, Bonnie Nardi, and Bill Tomlinson. 2017. Political Realities of Digital Communication: The Limits of Value from Digital Messages to Members of the US Congress. In *Proceedings of the 2017 Workshop on Computing Within Limits* (LIMITS '17), 129–138. https://doi.org/10.1145/3080556.3080565
60. Bonnie Nardi and Hamid Ekbia. 2017. Developing a political economy perspective for sustainable HCI. In *Digital Technology and Sustainability: Embracing the Paradox*.
61. National Research Council. 2012. *Computing Research for Sustainability*. The National Academies Press, Washington, DC, USA. https://doi.org/10.17226/13415
62. Jakob Nielsen. 1994. *Usability engineering*. Morgan Kaufmann Publishers, San Francisco, Calif.
63. William Odom, James Pierce, Erik Stolterman, and Eli Blevis. 2009. Understanding why we preserve some things and discard others in the context of interaction design. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems* (CHI '09), 1053–1062. https://doi.org/10.1145/1518701.1518862
64. Antti Oulasvirta and Kasper Hornbæk. 2016. HCI Research As Problem-Solving. In *Proceedings of the 2016 CHI Conference on Human Factors in Computing Systems* (CHI '16), 4956–4967. https://doi.org/10.1145/2858036.2858283
65. Daniel Pargman, Elina Eriksson, Mattias Höjer, Ulrika Gunnarsson Östling, and Luciane Aguiar Borges. 2017. The (Un)Sustainability of Imagined Future Information Societies. In *Proceedings of the 2017 CHI Conference on Human Factors in Computing Systems* (CHI '17), 773–785. https://doi.org/10.1145/3025453.3025858
66. Daniel Pargman and Barath Raghavan. 2014. Rethinking Sustainability in Computing: From Buzzword to Non-negotiable Limits. In *Proceedings of the 8th Nordic Conference on Human-Computer Interaction: Fun, Fast, Foundational* (NordiCHI '14), 638–647. https://doi.org/10.1145/2639189.2639228
67. James Pierce. 2012. Undesigning technology: considering the negation of design by design. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems* (CHI '12), 957–966. https://doi.org/10.1145/2207676.2208540
68. James Pierce and Eric Paulos. 2012. Beyond energy monitors: interaction, energy, and emerging energy systems. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems* (CHI '12), 665–674. https://doi.org/10.1145/2207676.2207771
69. Karl R. Popper. 1962. *Conjectures and Refutations: The Growth of Scientific Knowledge*. Basic Books.
70. Nicolas Prat, Isabelle Comyn-Wattiau, and Jacky Akoka. 2014. Artifact Evaluation in Information Systems Design-Science Research. *PACIS 2014 Proceedings*. Retrieved from http://aisel.aisnet.org/pacis2014/23

71. Pré consultants b.v. 2017. SimaPro. *SimaPro*. Retrieved September 10, 2017 from https://simapro.com/
72. Jenny Preece, Yvonne Rogers, and Helen Sharp. 2002. *Interaction Design*. Wiley, New York, NY.
73. Chris Preist, Daniel Schien, and Eli Blevis. 2016. Understanding and Mitigating the Effects of Device and Cloud Service Design Decisions on the Environmental Footprint of Digital Infrastructure. In *Proceedings of the 2016 CHI Conference on Human Factors in Computing Systems* (CHI '16), 1324–1337. https://doi.org/10.1145/2858036.2858378
74. Jan Pries-Heje, Richard Baskerville, and John Venable. 2008. Strategies for Design Science Research Evaluation. *ECIS 2008 Proceedings*. Retrieved from http://aisel.aisnet.org/ecis2008/87
75. Christian Remy, Oliver Bates, Thomas Vanessa, and Elaine May Huang. 2017. The Limits of Evaluating Sustainability. In *Proceedings of the Third Workshop on Computing within Limits*.
76. Christian Remy, Silke Gegenbauer, and Elaine M. Huang. 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), 1305–1314. https://doi.org/10.1145/2702123.2702567
77. Christian Remy and Elaine M. Huang. 2015. Addressing the Obsolescence of End-User Devices: Approaches from the Field of Sustainable HCI. In *ICT Innovations for Sustainability*, Lorenz M. Hilty and Bernard Aebischer (eds.). Springer International Publishing, Switzerland, 474.
78. David J. Roedl and Erik Stolterman. 2013. Design Research at CHI and Its Applicability to Design Practice. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems* (CHI '13), 1951–1954. https://doi.org/10.1145/2470654.2466257
79. David Roedl, William Odom, and Eli Blevis. 2017. Three Principles of Sustainable Interaction Design , Revisited. In *Digital Technology and Sustainability: Embracing the Paradox*.
80. Yvonne Rogers. 2004. New theoretical approaches for human-computer interaction. *Annual Review of Information Science and Technology* 38, 1: 87–143. https://doi.org/10.1002/aris.1440380103
81. Yvonne Rogers, Jenny Preece, and Helen Sharp. 2011. *Interaction design*. Wiley, Hoboken, N.J.; Chichester.
82. Antti Salovaara, Antti Oulasvirta, and Giulio Jacucci. 2017. Evaluation of Prototypes and the Problem of Possible Futures. In *Proceedings of the 2017 CHI Conference on Human Factors in Computing Systems* (CHI '17), 2064–2077. https://doi.org/10.1145/3025453.3025658
83. Phoebe Sengers and Bill Gaver. 2006. Staying Open to Interpretation: Engaging Multiple Meanings in Design and Evaluation. In *Proceedings of the 6th Conference on Designing Interactive Systems* (DIS '06), 99–108. https://doi.org/10.1145/1142405.1142422
84. Ben Shneiderman, Catherine Plaisant, Maxine Cohen, and Steven Jacobs. 2009. *Designing the User Interface: Strategies for Effective Human-Computer Interaction*. Pearson, Boston.
85. Elizabeth Shove, Mika Pantzar, and Matt Watson. 2012. *The Dynamics of Social Practice: Everyday Life and how it Changes*. SAGE Publications Ltd, Los Angeles.
86. M. Six Silberman, Lisa Nathan, Bran Knowles, Roy Bendor, Adrian Clear, Maria Håkansson, Tawanna Dillahunt, and Jennifer Mankoff. 2014. Next steps for sustainable HCI. *interactions* 21, 5: 66–69. https://doi.org/10.1145/2651820
87. M. Six Silberman and Bill Tomlinson. 2010. Toward an ecological sensibility: tools for evaluating sustainable HCI. In *CHI '10 Extended Abstracts on Human Factors in Computing Systems* (CHI EA '10), 3469–3474. https://doi.org/10.1145/1753846.1754003
88. Bruce Sterling. 2009. COVER STORY: Design Fiction. *interactions* 16, 3: 20–24. https://doi.org/10.1145/1516016.1516021
89. Erik Stolterman. 2008. The Nature of Design Practice and Implications for Interaction Design Research. *International Journal of Dsign* 2, 1: 55–65.
90. Alistair Sutcliffe. 2000. On the Effective Use and Reuse of HCI Knowledge. *ACM Trans. Comput.-Hum. Interact.* 7, 2: 197–221. https://doi.org/10.1145/353485.353488
91. Vanessa Thomas. 2017. Negotiating and engaging with environmental public policy at different scales. In *Digital Technology and Sustainability: Embracing the Paradox*.
92. Vanessa Thomas, Christian Remy, Mike Hazas, and Oliver Bates. 2017. HCI and Environmental Public Policy: Opportunities for Engagement. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*.
93. Kentaro Toyama. 2015. Preliminary thoughts on a taxonomy of value for sustainable computing. *First Monday* 20, 8. Retrieved August 25, 2017 from http://ojphi.org/ojs/index.php/fm/article/view/6132
94. United Nations. 2015. Paris Agreement.
95. United Nations. 2017. Sustainable development goals. *United Nations Sustainable Development*. Retrieved September 7, 2017 from http://www.un.org/sustainabledevelopment/sustainable-development-goals/
96. John Venable, Jan Pries-Heje, and Richard Baskerville. 2012. A Comprehensive Framework for Evaluation in Design Science Research. In *Proceedings of the 7th International Conference on Design Science Research in Information Systems: Advances in Theory and Practice* (DESRIST'12), 423–438. https://doi.org/10.1007/978-3-642-29863-9_31

97. John Venable, Jan Pries-Heje, and Richard Baskerville. 2016. FEDS: a Framework for Evaluation in Design Science Research. *European Journal of Information Systems* 25, 1: 77–89. https://doi.org/10.1057/ejis.2014.36
98. Martin Woolley. 2003. Choreographing obsolescence - ecodesign: the pleasure/dissatisfaction cycle. In *Proceedings of the 2003 international conference on Designing pleasurable products and interfaces* (DPPI '03), 77–81. https://doi.org/10.1145/782896.782916