interplay2015
DESIGN ▶ SCIENCE ▶ TECHNOLOGY ▶ ARTS

IASDR 2015 Congress
2-5 November 2015
Brisbane, Australia
http://www.iasdr2015.com

iasdr QUT

On What Grounds? An intra-disciplinary account of evaluation in research through design

Thomas Markussen, Aarhus School of Architecture, thomas.markussen@aarch.dk
Peter Gall Krogh, Aarhus School of Archtecture, pgk@aarch.dk
Anne Louise Bang, Kolding Design School. alb@dskd.dk

Abstract (Abstract heading style)
Research through design is a murky field and there is an increasing interest in understanding its varied practices and methodology. In the research literature that is initially reviewed in this paper two positions are located as the most dominant representing opposite opinions concerning the nature of such a methodology. One position proposes a cross-disciplinary perspective where research through design is based on models and standards borrowed from natural science, social sciences, humanities and art, while the other position claims a unique epistemology for research through design insisting on its particularities and warning against importing standards from these other disciplines. In this paper we argue for taking a third position, an intra-disciplinary position that appreciate how design processes and the making of artifact can be a method of inquiry, while at the same time insisting on using standards and terminology that can foster a dialogue with surrounding scientific cultures. To substantiate our claim we further introduce five methods of evaluation in research through design, which are derived from a close examination of a sample of PhD theses that are claimed to be exemplary of the field. In so doing, we aim to lay new grounds for a methodology.

Keywords: *Research through design; methodology; evaluation; interaction design; theory of science*

Introduction

Methodologically, research through design is a murky field. Asking practitioners and researchers about their notion of research through design, and you'll get an inconsistent set of methods and criteria defining the approach. This methodological plurality is mirrored in

the lack of standards for evaluation and agreed forms of output. If one looks into a sample of PhD theses that are claimed to be "exemplary" of the field, the huge and at times even irreconcilable diversity in evaluation practices is puzzling. At one end of the spectrum, evaluation is practiced as systematically and rigorous as in controlled lab settings (Frens, 2006; Ross, 2008), while at the other end research through design is dedicated more to activist reforming of practice and the crafting of manifestos than to evaluation (Trotto, 2011; Von Busch, 2008).

This "state of the art" has led some researchers to call for a policing of the research through design label, working out a formalized approach with an agreed upon method to document knowledge (Zimmerman, Stolterman, & Forlizzi, 2010). Other researchers, however, argue for appreciating the controversies and proliferation of research programs currently characterizing the field (Gaver, 2012). In caricature it can be noted that representatives of the first group works to associate design with changing existing research traditions (natural, technical, social sciences and humanities) dependent on the deployed methodology and measures for evaluation whereas the latter works to position design outside classical research and science. Interestingly we agree with both insofar as design delivers other results than classical sciences and thus exist outside such measures, but design research needs some common ground to discuss what is knowledge production and thus it needs a scientific foundation. However, such foundation does not derive from other fields identifying what is right and wrong, but we need to articulate design as an independent field of research that follows the same language games as any other science and research traditions articulating what we explore by means of hypothesising, experimenting, posing research questions and do evaluations.

We want to appreciate the rich diversity of evaluation practices and criteria that is characteristic of research through design. Yet, at the same time, we argue that research through design must still be judged against some kind of criteria of accountability (Koskinen & Krogh 2015 (in press) or validity of research (Hamilton & Jaaniste, udateret; Niedderer & Roworth-Stokes, 2007, s. 5). Diversity in evaluation does not preclude a coherent formal account of this diversity.

In order to get a firmer grasp of what evaluation means, we will go through some recent research literature that has taken up the question of evaluation in research through design. It is common in design research to cast light on evaluation by viewing it cross-disciplinarily from the perspective of science, social sciences and the arts and humanities. We shall argue, however, that there is a need to pay greater attention to how evaluation is actually being practiced within design research itself. Hence, we contend that an *intra-disciplinary* treatment looking at evaluation from the inside out is likely to be a valuable supplement to a cross-disciplinary understanding.

Having positioned our own work in relation to existing discussions, we will then examine a sample of PhD theses thereby identifying a set of evaluation practices. More specifically, we will demonstrate that evaluation in research through design can take the form of what we

shall refer to as *repercussive, relational, serial, expansive and eclectic* evaluation. It is our hope that these five methods of evaluation can be helpful to doctoral students, supervisors and researchers trying to navigate the expanding territory of research through design.

Related work

Research through design (hereafter abbreviated as RtD) has matured into an established research approach with a rapidly growing body of literature dealing with all sorts of foundational issues and questions for the discipline. The seminal work of Frayling (1993), Archer (1995) and Cross (2001) were vital in positioning research through design as a scientific culture of its own, a "third culture", so to speak, if we consider it existing next to the two other cultures suggested by C. P. Snow (1960): science and art. These epistemological distinctions – or "ways of knowing" (Cross) - are still lurking behind recent discussions, while the image has certainly become more nuanced. What currently divides researchers is the question whether one should insist on the epistemological autonomy of research through design or allow for various cross-disciplinary interplays. For instance, Koskinen et al. (2011) tend to treat research through design as if moulded by norms, methods and practices found within the natural sciences, the social sciences and art, while Stolterman (2008), Gaver (2012) and Bowers (2012) argue that RtD deals with criteria or topics that are unique – or "ultimate particulars" in Stolerman's terms – and therefore this research practice is irreducible to existing scientific models.

The question of evaluation hinges upon which one of these two epistemological positions one forfeits. Epistemology, to make sure, is concerned with *how* knowledge is gained by the use of a research method, and *why* results are obtained. According to Ziman (2002, s. 93), being able to give such an explanation, is the very essence of doing research. The *why*-question has to do with the disciplinary and personal motivations of the researcher (Bang, Krogh, Ludvigsen, & Markussen, 2012) and fundamental knowledge interests of his or her research community (something that can be compared with what Lakatos (1974) calls as a "research program"). To answer the *how*-question we need to work out viable accounts of how "means of design" (Fallman, 2005) and the making of artefacts can be considered a legitimate method of inquiry (Zimmerman et al., 2010). To be sure, it is a matter of explaining, not what methods are practiced by, let's say, the graphic designer, the interaction designer, the fashion designer, and so on, but rather how results can be documented and evaluated as an outcome not being reducible to the design work alone.

Methodologically, RtD is said to be "constructive" (Koskinen et al., 2011). That means, it is a research practice that changes the part of reality that is the very subject matter of the research, and the design researcher is taking active part in that process of change. Originally, Simon (1969) suggested that this change consisted in transforming reality into a "preferred state". But as Dunne and Raby (1999; 2001) have abundantly made clear what is deemed "preferable" is not a reliable evaluation criteria. Indeed by introducing terms such as

"inhuman factors" and "unfriendliness" as design ideals Dunne and Raby encourage us to reflect on the unspoken norms, politics, ideologies and cultures that permeate any act of designing and design research, but which are rarely highlighted.

These ideologies, politics and cultures must be taken into account, if one wants to understand how the outcome of design and design research is evaluated. Koskinen (2015) argues that there are four dominating "cultures of analysis" in design research, each of which prescribes certain ways of assessing and evaluating results. The *first* culture largely derives its model of analysis from the natural sciences and psychology. Here evaluation is arrived at through statistical analyses of data, which is evaluated according to their fit with a hypothesis formulated on the basis of theory (ibid., p. 218).

The *second* culture is influenced by the social sciences. Here evaluation is typically relying on the use of "analytic induction", where a hypothesis is not based on theory, but created out of an analysis of a small number of cases and empirical data gathered from fieldwork. Such an analysis typically takes the form of post-it clustering or the analysis of design probes rather than statistical analysis. The unit of analysis is not statistical data and measures as in the first culture, but 'meaning' and 'context' (p. 219).

The *third* culture borrows its model of analysis from the humanities. Basically, says Koskinen, this model can be characterized as 'explanation', which is further defined as "a detailed examination of meaning" governed by the methodological principles of the hermeneutic circle (p. 221). The hermeneutic circle represents the idea that every act of explanation – and consequently of evaluation – must be seen as a movement back and forth between understanding parts of a phenomena in relation to a larger whole[1]. This explanation does not come in the form of scientific explanation, but as cultural understanding influenced by the individual's common sense, preconceptions, taste and past biography (cf. Moran, 2000, s. 280).

The *fourth* analytic culture is art and design-based. According to Koskinen this means that scientific ideals such as "transparency", "clarity" and "disinterested knowledge" is given up in favour of "idiosyncrasies and vague analysis; in fact, ambiguity may even be encouraged, if it leads to interesting design." (Koskinen, 2015, s. 222). Evaluation here hinges entirely on the subjective judgement of design qualities and so-called "creative steps" in the design process, and less on whether other design researchers can understand how design work is produced.

While we generally agree with Koskinen in that design research has borrowed research practices and methods "from disciplines with longer historical roots" (p. 217), we also find that his account is not entirely satisfying. Koskinen is of the belief that "the best way to understand analysis in design research is to look at it from an *abstract perspective*" (p. 224, our italics). We, on the other hand, argue that by taking such a perspective there is a risk of

[1] Cf. The Stanford Encyclopedia of Philosophy: http://plato.stanford.edu/entries/hermeneutics/

loosing sight of how design serves as a method of inquiry of its own. Koskinen's account is cross-disciplinary insofar as he treats analysis in design research as if it needs to justify itself by lending models from other scientific disciplines. Except, of course, from the art and design-based analytic culture. But, according to Koskinen, 'design-based' is synonymous with 'idiosyncratic', 'vague analysis', 'ambiguous' and 'definitely not scientific'; a kind of black art unable (or unwilling) to participate in the language game about what should be considered reliable and valid criteria of sound research.

Rather than a cross-disciplinary approach designed on the basis of adopting a variety of methods from traditional research disciplines, we argue for the necessity of developing an *intra-disciplinary* account of how evaluation is practiced in RtD. By intra-disciplinary we understand an account that takes a look at evaluation from the inside out – as it is developed in a research project out of a series of design experiments executed in order to cast light on a research question or hypothesis (Bang et al., 2012). Interestingly, an attempt at such an account is made by Zimmerman et al. (2010). Lamenting that "there is no agreed upon method to document knowledge […] that emerge from research through design", the authors suggest a set of formal distinctions enabling design researchers to classify various types of theory that may be build up from RtD. More specifically, such theory may take the form of *guiding philosophies, conceptual frameworks, implications for design and design implications*. What characterizes all of these outcomes is that they are typically developed with the intention of improving design practice. Hence, they serve as theories *for* design. For instance, guiding philosophies sensitize "concepts to help direct designers and researchers in solving design problems"; implications for design results "from inquiry into wicked problems"; and design implications arise "from the analysis of designed artefacts" (Zimmerman et al., 2010 p. 313).

Zimmerman and his colleagues provide a valuable formal account that enables researchers to classify research outcomes into certain types of theories for design. Yet, the relationship between such theories and design work needs to be further specified. On what grounds do we judge whether a theory for design is useful, valuable or successful? What is the validity and role of theory produced from design?

Addressing these questions, Gaver (2012) points out that we should not evaluate theory produced from design on the same conditions as in science. The evaluation criteria are simply too different. A theory in RtD is not verifiable through falsification (p. 943). Its role is to be 'generative and suggestive' (ibid. 943) rather than to explain universal truths and principles of empirical facts. More interestingly, however, Gaver observes "that theory underspecifies design […] in the sense that many aspects of a successful design will not be captured by a given theory." (Gaver, 2012 p. 944). Hence, if we focus too rigidly on theory as the sole outcome of RtD, we neglect that design work is the fundamental achievement. Design work does not serve as mere illustration or exemplification of theory. On the contrary, design takes centre stage, while theory serves the function of annotating resulting artefacts, interfaces and products, making visible certain features, rationales and choices made by the

designer. As in the portfolio of Dieter Rams where theory is manifested in the form of Ram's well-known 10 concise design principles (pp. 944-45). This is a more accurate specification of what we should expect from RtD: *annotated portfolios*.

Under the heading "The Logic of Annotated Portfolios" Bowers (2012) provides further insight into how annotated portfolios are constituted and what role annotations fulfill. Annotations are what make a collection of designed artifacts into a portfolio. Annotations bring together artifacts as 'a systematic body of work'; explicate 'family resemblances and differences' between works within the portfolio or in relation to related work in a field. Annotations can also 'configure use, appreciation, aesthetics and scientific value as well as suggesting future research and design possibilities' (p. 72). In this way design works "are organized, categorized and otherwise arranged in their presentation to have or illustrate a point or several points and through doing this to reveal … something of the design identities in the work and the nature of the *contribution* being made" (p. 71).

Following Gaver's and Bowers' account, we argue that by looking more closely into how portfolios are annotated, it is possible to get a firmer grasp of how evaluation is practiced in RtD, because one of the primary tasks of an evaluation is to reveal "the contribution being made". Furthermore, we suggest conceiving of a selection of PhD thesis as exemplars of annotated portfolios each of which is constituted by certain *methods of evaluation*. More specifically, we will demonstrate that these methods can be characterized appropriately as being *repercussive, relational, serial, expansive or eclectic*.

Five Methods of Evaluation in Research through Design

In this section we will look closer into six PhD theses that have been selected because they are representative of varied ways in which evaluation in RtD is being practiced (without claiming our treatment to be exhaustive). Further the theses broadly cover design practice from fashion and product design to interaction design. These PhD theses served as the curriculum for three doctoral courses dedicated to understanding basic methodological principles of RtD. Over 60 PhD students from universities and design schools all over Europe and Scandinavia participated in the courses. Due to this contextual setting, regions and countries outside Europe and Scandinavia are unfortunately misrepresented.

This section is organized into five sub-sections each focusing on a particular logic of evaluation. In each sub-section we will provide general characteristics as well as examples taken from the selected theses.

Repercussive Evaluation

This method is generally characterized by the evaluation of experimental results and design work according to one nucleus of criteria. To secure as controlled an evaluation as possible all disturbing factors and contextual relationships are excluded. All insights gained through

sketching, mock-ups, and prototyping during the design process are then held up strictly against to the nucleus of criteria. As design iterations are performed knowledge outcomes may layer themselves as circuits around the nucleus, but the understanding of every outcome is gained only by falling back into the center, which is why we refer to it as repercussive (see Figure. 1).

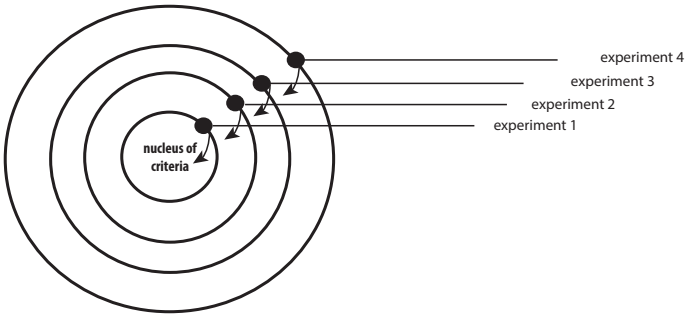

Figure. 1 – repercussive evaluation

The method of repercussive evaluation can be found in the work of Frens (2006) and Ross (2008). Ross, for instance, uses a research through design cycle to investigate how three different perspectives on interaction behavior (Dynamic Form, Social Activity and Sensory-Motor Activity) can inform the design of intelligent lamps. Furthermore, he wants to investigate whether it is possible in the design of the lamps to create conditions so as to elicit certain values in user's experiences of the lamps. For his lamp design, Ross deliberately chose neutral colors, material and a static shape to rule out any influence from factors other than those of interest. Through a series of experience prototyping experiments documented through microanalysis, Ross explores various ways of correlating dynamic form, social activity and sensory-motor levels. While this results in detailed design guidelines for how the three perspectives may inform design of intelligent lamps, his evaluation shows that the lamp designs do not elicit the right values in user experience. Hence, Ross ends up falsifying one his hypotheses.

Relational Evaluation

Relational evaluation is at stake when evaluation is used primarily to explicate and judge how design work exists in a system of family resemblances and differences. Similarities and differences are relations that can be either *intrinsic* as in between individual artifacts in a designer's portfolio; or the relations can be *external* as when design work is compared with the works of other designers. As a requirement for this method of evaluation, a core criterion must be established (e.g. a key concept, term, measure or theme), which can serve as grounds of comparison of how design work relate intrinsically and externally (see Figure 2). Even though relational evaluation seems at first sight to be indistinguishable from a classical

comparative study, it appears under closer scrutiny that comparative evaluation accounts only partly for the method of relational evaluation.

This method is widely used and well documented in the thesis by Kinch (2014) and Niederer for example. Kinch evaluates how people perceive and use an interactive bench in three different contexts (an airport, a concert hall and a shopping mall). A key evaluation criterion for her is 'atmosphere experience', which is a hybrid concept founded upon theories of user experience as well as on philosophical speculations on the nature of atmosphere. By studying people's experiences of one and same bench as it is contextually replaced, Kinch is able to uncover several dimensions of atmospheric experiences. The evaluation is first and foremost concerned with intrinsic relations.

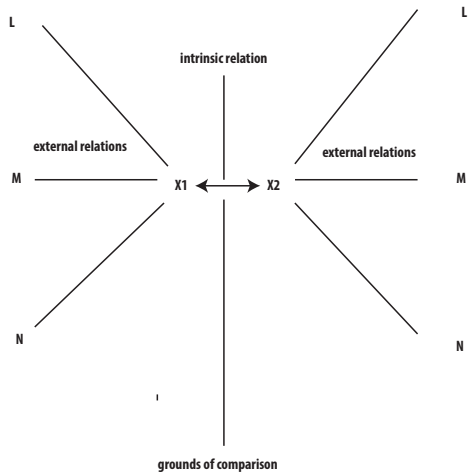

Figure 2 - relational evaluation. The model can be used to map the logic governing how design work is evaluated in a system of relations. In Niederer's case performative objects act as ground for comparison; L represents Art Objects; M represents Ritual Objects; N represents Craft Objects. X1 is the Libation Cup project, while X2 is the Social Cups projects. The internal relation is concerned with difference in terms of mindfulness, while the external relations indicate how the disruptions of function in performative objects differ from Art, Ritual and Craft Objects.

In Niederer both intrinsic and external relations figure in her evaluation. Niederer sets up the concept of 'performative objects' as a core criterion for conducting an evaluation of her design work. She draws upon several theories (e.g. semiotics, hermeneutic phenomenology, behaviorist theory) in order to derive two defining sub-criteria of performative objects: Thus, performative objects are objects that are designed to i) cause an experience of mindfulness (what Niederer refers to as 'result') through ii) a disruption of function (referred to as 'means'). Through a careful analysis, Niederer then identifies four object categories against which her performative objects are compared externally: *art objects, ritual objects, design objects and craft objects*. Art objects, ritual objects and craft object are able to evoke mindfulness of a similar kind as performative objects, but they do so through different means than disruption of function. Only design objects qualify in this respect.

However, Niederer's relational evaluation is not restricted only to the external comparisons between performative objects, on one hand and art, ritual and craft objects on the other. Taking center stage is her evaluation of how the framework she has set up allows her to explore some design possibilities for designing drinking vessels as performative objects. Hence, in the Libation Cup project Niederer deliberately designs a cup with wholes in it to evoke mindfulness-of-self in relation to the object and symbolic connotations, while, in the Social Cups project, cups without foots are designed to encourage users to be mindful of their "interpersonal interaction" during the drinking act as the cups can only be placed to stand on a table when combined in sets of three.

To sum up: Niederer uses evaluation of external relations to name and classify performative objects as a new object category, while her evaluation of intrinsic relations is central for making a nuanced distinction between two forms of mindfulness.

Serial Evaluation

The method of serial evaluation denotes how design experiments are being evaluated according to a certain order or logic of locality determined by how experiments in a sequence has cast light on an overall research interest. Like relational evaluation, knowledge production in the progressive method is achieved on the basis of inquiries into relationships between design experiments. But there is a significant difference. Whereas, in the first instance, insights from design experiments are held up in a system of relations (cf. Figure 2), in serial evaluation, it is the local relationship between two neighbouring experiments that matters (see Figure 3).

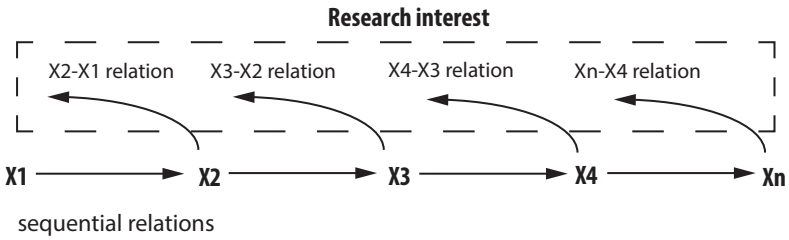

Figure 3 - serial evaluation. X represents design experiments performed in a sequence. Left pointing arrows stands for evaluation of local relationships between two neighbouring experiments.

In the work of Lynggaard (2012), for example, a set of so-called *tactics for making home* is derived from her ethno-methodological studies of highly mobile people. One tactic is Connecting – staying in touch with those at home over distance; another one is Spreading – distributing one's belongings in a space to make it homely. Rather than starting out from a

fixed nucleus of criteria as Ross, Lynggaard takes one tactic at a time exploring it further in concrete experiments. Thus, serially, each tactic experiment is evaluated according to the previous experiment, so that, in the end, Lynggaard is able to present a taxonomy consisting of seven homing tactics. This is also different from Niederer insofar as Lynggaard does not so much use theory to evaluate her design work. Rather Lynggaard uses design work to build new theory.

Expansive Evaluation

This method of evaluation focuses on how designerly experiments can serve to reveal and identify qualities of an area as-yet uncovered. It resembles a voyage of discovery where new places and insights are described along the way much like annotating bits of observation and information onto the traveller's map. Such is the case, for instance, in Dinder who introduces 'engagement' from the theories of notably Berleant, Borgman and pragmatism as a valuable new territory for designing 'participatory engaging' exhibitions in museums.

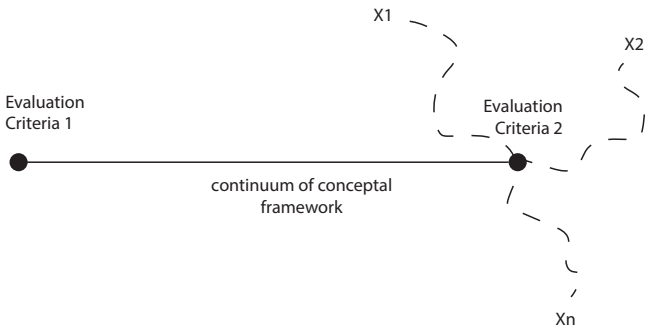

Figure 4 - expansive evaluation. In Dindler passive user involvement would be Evaluation Criteria 1, while active user involvement would correspond to Evaluation Criteria 2. The IXP prototype for Kattegat Marin Centre is represented by X1; the interactive runic stone for Moesgaard Museum is X2. The two prototypes show different aspects of active user involvement.

Dindler takes his point of departure in analyzing the fairly traditional museum space of the Viking Ship Museum in Roskilde, where active involvement of the audience lies at an absolute minimum. This serves as a low-level indicator for his evaluation of two interactive prototypes, which he developed and designed for the Kattegat Marine Centre and Moesgaard Museum respectively. In contrast to a control study evaluation, Dindler does not use the low level indicator as a kind of baseline measure for comparing user engagement in the two prototypes with the level of engagement in the Viking Museum. Rather, he regards minimum engagement in the traditional museum space as representing one end of a spectrum, while his two prototypes represent the opposite end: 'active user involvement'. At the same time, he is able to broaden what forms active user involvement may take by evaluating his prototypes.

In the Kattegat Marine Centre the installation allows for "engaging visitors in playful activity" in creating imaginative fish and sea creatures, whereas, in Moesgaard Museum, the possibility of creating their own interactive runic stones, prompt visitors to reflect on seminal events in their lives to be told to other visitors. In this way Dindler is capable of expanding the notion of engagement.

Eclective Evaluation

Eclective evaluation is characterized by its way of fusing and sampling ideas, theories and philosophy from different disciplines. This sampling is often driven by the peculiar interest and agenda of the researcher and can be overtly normative, ideological or political. Unlike relational evaluation, eclective evaluation is not focused on explaining how results fit into a coherent conceptual system. Nor how a philosophical notion can be made sensitive to designers as in Dindler. Rather, eclective evaluation typically takes the form of manifesto-like claims or calls for aesthetic reform of practice as documented by von Busch (2008) and Trotto (2011) for example.

In von Busch's thesis there is hardly any evaluation in the traditional sense. Von Busch is interested in reforming the fashion system from the bottom-up and democratizing it by making it possible for ordinary people and professional amateurs (so-called *proams*) to have a say in the shaping of fashion. One of his central arguments is that, to let this happen, a new anarchistic approach has to be invented, which enables collaboration between fashion designers and non-designers and which he refers to as 'hacktivism'.

Busch draws upon theories of DIY, software hacking and political theories to formulate the conceptual foundations of hacktivism and he sets up a series of open workshops and exhibitions to explore if people and proams are able to reform the fashion system. The workshops are not tied together by certain logic of progression or expansion, but exist as singular events (see Figure 5) planned to explore the fundamental question: Is it possible?

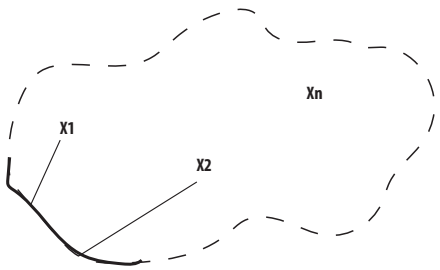

Figure 5 - eclective evaluation. Design experiments (W1, X2, Xn) are performed independent of each other and evaluation of their individual success or failure lead to demonstrating the possibility of an overall research aim.

In his evaluation of one of these workshops held during the Hackers and Couture Heretics Exhibition, Busch concludes that it has been a success for two reasons. First, it was easy for people and bystanders to participate due to the use of simple methods such as re-design, button exchange, shop dropping etc. (ibid., p. 219). Secondly, the workshop proved that "DIY does not have to look grubby or lack craftsmanship" (p. 222).

These observations tell a great deal about the nature of eclective evaluation. It rests entirely upon subjective judgments, and the purpose is first of all to highlight *that* people actually participated rather than *how* they did it. The ultimate goal for Busch is to demonstrate that hacktivism is possible and that it does not rule out aesthetic quality. Eclective evaluation is thus used by Busch, not to document the validity of results, but as means for articulating a new approach and aesthetics founded upon DIY, activist philosophy and political theory.

Discussion

On the basis of our description and exemplification of five methods of evaluation, we are able to account more accurately for how we contribute to existing research literature. First of all, we have attempted to show that an intra-disciplinary perspective on evaluation in RtD offers insight into concrete methods that is not captured by Koskinen's cross-disciplinary perspective. For instance, if we consider Koskinen's four cultures of analysis, then Kinch and Lynggaard would fall under the second culture, while Niederer and Dindler would belong to the third culture. However, by taking an intra-disciplinary approach, the kinship between the researchers stands out differently. Thus, both Kinch and Niederer perform a relational evaluation, while Dindler og Lynggaard conduct two separate forms of evaluation referred to as serial and expansive. This indicates that the cross-disciplinary framework is too coarse grained, and we suggest that an intra-disciplinary could serve as a valuable supplement in several respects.

Secondly, while we find the notion of annotated portfolios valuable, we also believe that it needs further elaboration. Both Gaver and Bowers argue that theory is useful for annotating design work in a portfolio, thereby making visible certain features, family resemblances and differences. Yet, by looking closer at methods of evaluation, we are able to explain in more depth how design work is organized according to certain evaluation criteria and for what purpose. More precisely, we have claimed that design work can be organized according to a nucleus of criteria (Frens, Ross), intrinsic and external relations in a system (Kinch, Niederer), serial relations between design work (Lynggaard), an identified concept un-theorized by an existing framework (Dindler) or isolated design experiments arranged for reforming practice (Busch, Trotto). We suggest that these five methods of evaluation correspond to what Bowers refers to as the logic of annotated portfolios, and that they are helpful for planning, performing and examining doctoral work. But further studies are needed before this can be ascertained.

Thirdly, even though our study thus contributes with valuable new knowledge on RtD methodology, its explanatory scope should not be overestimated. Our methodology is derived from the study of a few PhD theses and therefore future work is needed to determine if the description of evaluation methods is accurate and systematic. Such work would benefit from studying a large number of PhD theses representing and challenging the five methods. Moreover their appliance to doctoral work found at other continents must be investigated.

Conclusion

In this paper, we have argued that RtD would benefit from the working out of a methodology that appreciates the rich diversity of the field. In the research literature two positions are dominant representing opposite opinions concerning the criteria and purposes of such a methodology. One position proposes a cross-disciplinary perspective where RtD is based on models of analysis and evaluation borrowed from existing scientific cultures, while the other position claims a unique epistemology for RtD insisting on its particularities and warns against importing standards from other disciplines. We have argued for taking a third intra-disciplinary position that is premised on the idea that the cross-disciplinary position is insufficient because it cannot account for how design processes and the making of artifacts serve as methods of inquiry. At the same time we deem it relevant and critically important for the field to participate in the language games of science by using vocabulary and terminology that is familiar to surrounding research cultures. To substantiate this argument we introduced evaluation as a useful term and have looked carefully into five methods of evaluation in our examining of doctoral work.

This paper is the last out of three that address key issues and foundations of RtD. In our previous work we have focused on hypothesis-making and the crafting of research questions in RtD (Bang et al., 2012) as well as on the methods of experimentation in RtD (Krogh, Markussen, & Bang, 2015) It is our hope that this work will have a practical value for doctoral students who are planning and doing RtD, that it will be helpful for supervisors and examiners assessing doctoral work, and last but not least that it can foster a constructive dialogue and an interplay with research disciplines outside the design research community.

References

Archer, B. (1995). The Nature of Research, 6–13.

Bang, A. L., Krogh, P., Ludvigsen, M., & Markussen, T. (2012). The Role of Hypothesis in Constructive Design Research. I *The Art of Research IV*. Aalto Univeristy, Helsinki.

Bowers, J. (2012). The logic of annotated portfolios: communicating the value of'research through design'. I *Proceedings of the Designing Interactive Systems Conference* (s. 68–77). ACM.

Cross, N. (2001). Designerly Ways of Knowing: Design Discipline Versus Design Science, *17*(3), 49–55.

Dunne, A. (1999). *Hertzian tales: electronic products, aesthetic experience, and critical design*. London: RCA CRD research publications.

Dunne, A., & Raby, F. (2001). *Design noir: The Secret Life of Electronic Objects*. Basel: Birkhäuser.

Fallman, D. (2005). research-oriented design isn't designoriented research. I *Proceedings of Nordes: Nordic Design Research Conference*. Citeseer.

Frayling, C. (1993). Research in Art and Design, *1*(1), 1–5.

Frens, J. W. (2006). Designing for rich interaction: Integrating form, interaction, and function. Eindhoven University of Technology

Gaver, W. (2012). What should we expect from research through design? I *Proceedings of the SIGCHI conference on human factors in computing systems* (s. 937–946). ACM.

Hamilton, J., & Jaaniste, L. (udateret). The Effective and the Evocative: Practice-led Research Approaches Across Art and Design. Montreal.

Kinch, S. (2014). *Designing for Atmospheric Experiences. Taking an Architectural Approach to Interaction Design* (Ph.D. Dissertation). Aarhus School of Architecture, Aarhus.

Koskinen, I. (2015). Four Cultures of Analysis in Design Research. I *The Routledge Companion to Design Research* (s. 217–224). Oxton: Routledge.

Koskinen, I., Zimmerman, J., Binder, T., Redström, J., & Wensveen, S. (2011). *Design research through practice*. Amsterdam: Morgan Kaufmann.

Krogh, P. G., Markussen, T., & Bang, A. (2015). Ways of Drifting - 5 Methods of Experimentation in Research through Design. Præsenteret ved ICord'15 - International Conference on Research into Design, Indian Institute of Science, Bangalore: Springer-Verlag.

Lakatos, I. (1974). The role of crucial experiments in science. *Studies in History and the Philosophy of Science*, *4*(4).

Lynggaard, A. (2012). *Homing Interactions: Tactics and Concepts for Highly Mobile People.* Aarhus School of Architecture, Aarhus.

Moran, D. (2000). *Introduction to phenomenology*. London: Routledge.

Niedderer, K., & Roworth-Stokes, S. (2007). The role and use of creative practice in research and its contribution to knowledge. I *IASDR International Conference*.

Ross, P. (2008). *Ethics and Aesthetics in Intelligent Product and System Design*. Eindhoven University of Technology, Eindhoven.

Simon, H. A. (1969). *The sciences of the artificial*. The MIT Press.

Snow, C. P., & Snow, B. (1960). *The two cultures and the scientific revolution*. Cambridge University Press New York.

Stolterman, E. (2008). The nature of design practice and implications for interaction design research. *International Journal of Design*, *2*(1), 55–65.

Trotto, A. (2011). *Rights through Making – skills for pervasive ethics*. Eindhoven University of Technology, Eindhoven.

Von Busch, O. (2008). *Fashion-able. Hacktivism and engaged fashion design [Elektronisk resurs]*. University of Gothenburg. Hentet fra http://hdl.handle.net/2077/17941

Ziman, J. (2002). *Real science: What it is and what it means*. Cambridge University Press.

Zimmerman, J., Stolterman, E., & Forlizzi, J. (2010). An Analysis and Critique of Research Through Design: towards a formalization of a research approach. I *Proceedings of DIS 2010*. Aarhus: ACM.