The impact of concept (re)presentation on users' evaluation and perception

Sarah Diefenbach, Marc Hassenzahl, Kai Eckoldt, Matthias Laschke
Experience Design and Ergonomics
Folkwang University of Arts, Essen, Germany
{sarah.diefenbach, marc.hassenzahl, kai.eckoldt, matthias.laschke}@folkwang-uni.de

ABSTRACT
Early product concept evaluation, which is based on descriptions or conceptual sketches instead of functional prototypes or design models, has many practical advantages. However, a question at hand is whether the format of representation impacts the results of empirical "user studies". A study with two different design concepts and 326 participants revealed that global product evaluation (i.e., goodness) and high-level product perceptions (i.e., pragmatic quality, hedonic quality) are not influenced by differences in the concept (re)presentation (text, pictures, video, functional prototype). Only the assessment of interaction characteristics, such as its speed, was affected.

Author Keywords
Prototyping, concept testing, representation, evaluation, user experience, interaction,

ACM Classification Keywords
H5.2. Information interfaces and presentation (e.g., HCI): User interfaces -- Evaluation/methodology

INTRODUCTION

A major challenge for product designers and industry is the early identification of product concepts, which will be understood, accepted, and at best "loved" by potential future customers. It is common practice to decide on promising and less promising ideas by means of early "user studies." At that stage, however, concepts are available rather as textual descriptions or rough conceptual sketches than functional prototypes or design models. The benefit of "testing" descriptions or sketches early on is obvious. It demands fewer resources and, thus, allows for the empirical exploration of a greater number of alternative concepts or more revisions over time. However, it is an open question, whether concept evaluations gathered from potential users are biased by the way the concept is (re)presented. Intuitively, one could assume that users require first-hand interaction with a concept to provide a valid evaluation. At least, one would expect differences in the assessment of the same concept, depending on the way it was (re)presented.

NordiCHI 2010, October 16–20, 2010, Reykjavik, Iceland.
 ISBN: 978-1-60558-934-3...$5.00.

In Human-Computer Interaction (HCI) research, this topic has rarely been studied systematically. The general debate on low versus high fidelity prototypes is nothing new [e.g., 6], but so far, this discussion aimed mainly at costs and practical benefits rather than the comparability of results. There have been some valuable comparative studies in the recent years [e.g., 4, 5, 7]. However, they rely on contrasting selective kinds of prototypes, such as video and storyboard format [e.g., 7] or paper, computer, and fully operational prototypes [e.g., 4, 5], often with a focus on usability measures, and only one tested concept.

The aim of the present study is a further exploration of representation format, which enhances previous research in two ways: First, we were interested in whether there is a systematic effect of the "richness" of representation on different aspects of concept evaluation and perception. We used the probably simplest form of representation, i.e., textual description, as a baseline, which was then systematically enriched by additional information, resulting in four representation conditions. Second, any potential effect of different representations should be checked for stability. To do so, we studied two different design concepts at once. This allowed for testing potential interaction effects, which would indicate an unsystematic and unpredictable effect of the format of representation, and thus a continuous risk of bias in evaluation studies. In addition, studying two different concepts allowed for a check of the general reactivity of measures, since we expected to detect differences between the two concepts.

In the following, we present a study of two product concepts, each (re)presented in four different ways.

STUDY

Three-hundred twenty-six individuals (215 female, mean age=35, SD=12.81, min=15, max=70) participated in the study. They were randomly assigned to assess one of two concepts, each presented in one of four different representation conditions. Consequently, the study had a

4x2 between-subjects design, with representation and concept as independent variables.

The two product concepts

Both concepts suggested a novel and unusual way of interacting with a lamp. The *candle lamp* looked like a standard lamp, hanging above a dining table (Figure 1). Switching it on required to "light" an attached, wick-like element (actually a heat sensor) with a match or a lighter. It was switched off by blowing against the shade. *Forget-me-not*, the second concept, resembled a flower. In the moment a "petal" is touched, *forget-me-not* opens and lights up. However, the petals start to close again, which dims the light. To get light over time, thus, requires the user to touch the lamp repeatedly. The objective of both concepts was to make the interaction with the lamps and the according energy-consumption more conscious, i.e., to re-emphasis the value of energy.

(Re)presentations

Inspired by media richness theory, which considers documents as the leanest and face-to-face contact as the richest form of communication [1], we studied an according continuum of (re)presentations, reaching from textual description to first-hand interaction:

Text (Baseline). Participants got a short text describing the concept. For example, the *candle lamp* was described as follows: "The operation of the lamp is based on the candle metaphor. The lamp is switched on by briefly heating a wick-like element (attached to the lamp shade) with a lighter. Blowing into a circular opening in the middle of the shade switches off the light. The concept aims at making energy consumption more conscious."

Text and pictures. In this condition, a sequence of four pictures, which demonstrated the single steps of operating the concept, was added to the textual description (see Figure 1). The photos showed a real person interacting with the concept in a real environment. However, the product was only roughly sketched and was added subsequently, producing the impression of a montage.

Text and video. In addition to the descriptive text, there was a short video of a person interacting with a prototype and thereby demonstrating the operation of the lamp (see figure 2 for sample screenshots).

Text and real interaction. In addition to the descriptive text, participants experienced the interaction themselves by means of a functional prototype (the same as used in the *video* condition).

Note that the descriptive text served as a "baseline of information" in all the four representations. The crucial factor between the conditions was the richness of *additional* information. Significant differences between judgments based on leaner and richer formats of concept representation would indicate that the former were not sufficient to get the product idea across – at least regarding certain aspects. In order to identify such differential effects of the type of representation on different assessments we intentionally addressed different levels of product evaluation and perception.

Figure 1: Pictures representing the *candle lamp* (left) and *forget-me-not* (right).

Figure 2: Stills from the video representing the *candle lamp* (left) and *forget-me-not* (right).

Measures

After the presentation of the concept, participants were asked for a number of assessments common in user studies: global product evaluation, perceptions of the product character (i.e., pragmatic, hedonic), and interaction characteristics (e.g., *slow–fast*). The global product evaluation was measured with a single seven-point semantic differential item, capturing a product's general "goodness" (i.e., *bad–good*) [3]. Such global, high-level evaluations might be the only basis for decisions on which concept seems worthwhile, especially at an early conceptual stage, with a high number of alternatives. The perceived product character was assessed with an abridged version of the AttrakDiff2-questionnaire [3]. It consists of eight seven-point semantic differential items, four measuring task-related, pragmatic quality (e.g., *simple–complicated*, Cronbach's Alpha=.74), and four self-related, hedonic quality (e.g., *dull–captivating*, Cronbach's Alpha=.82). Scale values were computed by averaging the according item values (the inter-scale correlation was .42, which is substantial, but still considerably smaller than the Alpha's). To capture the participants' perception of interaction characteristics, we used the *Interaction Vocabulary* [2]. The vocabulary consists of eleven dimensions capturing aspects of the perceived aesthetics of interaction: *speed, power, continuity, precision, directedness, spatial proximity, immediacy, change, delay, evidence,* and *need for attention*. Each dimension is represented by a single seven-point semantic differential item with according verbal anchors, such as *slow–fast*, *gentle–powerful*, or *approximate–precise*. All materials were in German.

While the global evaluation may easily be based on a mere description of a product concept, the perception of whether an interaction with a product – even if well-described – is gentle or powerful, precise or proximate may be difficult without hands-on experience. Though the study was essentially exploratory, we expected the effect of representation on assessments to be stronger on the level of interaction (i.e., captured by the *Interaction Vocabulary*) than on the level of global evaluation and product perception.

RESULTS AND DISCUSSION

Goodness

A 4x2 analysis of variance with *representation* (text, text and pictures, text and video, text and real interaction) and *concept* (candle lamp, forget-me-not) as independent and *goodness* as dependent variable revealed no significant effect. Neither did *representation* take effect on how good or bad participants rated the concept, $F(3, 318)=1.05$, $p>.05$, nor was one concept rated as better than the other, $F(1, 318)=0.39$, $p>.05$. The interaction effect remained insignificant. A t-test comparing the overall mean *goodness* ($M=4.49$, $SD=1.63$) against the mid-point of the scale (4) showed that the concepts were valued rather positive than negative, $t(325)=5.89$, $p < .001$. The textual description was obviously sufficient to convey the general idea. Additional information (pictures, video, real interaction) did not impact the global product evaluation.

Pragmatic and hedonic quality perception

Two separate 4x2 analyses of variance with *representation* and *concept* as independent and either *pragmatic quality* or *hedonic quality* as dependent variables revealed no significant main effect for *representation* (*pragmatic quality*, $F(3, 318)=1.44$, $p>.05$; *hedonic quality*: $F(3, 318)=1.85$, $p>.05$). However, there was a main effect of the *concept*, $F(1, 318)=19.80$, $p<.001$: *forget-me-not* was assigned a significantly higher degree of pragmatic quality than the *candle lamp*, see Figure 3 for means in the four representation conditions. *Forget-me-not* was also perceived as more hedonic $F(1, 318)=3.94$, $p=.048$, see Figure 4 for mean values in the four representation conditions. However, the difference between concepts was less pronounced for hedonic compared to pragmatic quality. No interaction effect was found. Hence, no matter whether participants' assessments were based on real interaction, seeing a video, a picture story, or reading the concept description only, they all came more or less to the same conclusion.

Interaction vocabulary

Eleven separate 4x2 analyses of variance with *representation* and *concept* as independent and the 11 dimensions of the interaction vocabulary as dependent variables revealed significant main effects for *representation* on the dimension *speed* (i.e., *slow–fast*), $F(3, 318)=3.89$, $p<.01$, and *change* (i.e., *stable–changing*), $F(3, 318)=3.30$, $p<.05$. The pairwise post-hoc comparison of group means (Scheffé test) showed that the interaction felt significantly faster (*real interaction* condition: $M=4.47$, $SD=1.59$) than it had been expected on the basis of the textual description (*text* condition: $M=3.75$, $SD=1.75$) or the static pictures (*pictures* condition: $M=3.65$, $SD=1.76$). Assessments based on seeing a video of someone interacting with the product (*video* condition: $M=3.79$, $SD=1.69$) converged to those based on first-hand interaction, the difference failed to reach statistical significance. Regarding *change*, only one significant difference between representations emerged: the interaction was perceived as more stable in the *real interaction* condition ($M=3.46$, $SD=1.89$) compared to the picture condition ($M=4.37$, $SD=1.72$).

Besides *representation*, also a main effect of *concept* on *speed*, $F(1, 318)=8.83$; $p<.01$, and *change*, $F(1, 318=19.50$; $p<.001$) was found. Interacting with *forget-me-not* was assessed as fast ($M=4.19$, $SD=1.68$) and changing ($M=4.39$, $SD=1.85$) whereas interacting with the *candle lamp* was assessed as slow ($M=3.66$, $SD=1.74$) and stable ($M=3.58$, $SD=1.66$). Those differences are quite obvious. The action of "lighting up" and "blowing out" the *candle lamp* takes time and, thus, appears as rather slow compared to the touch of the petals of *forget-me-not* and their immediate reaction. Moreover, while the light of *forget-me-not* is dimmed automatically and, thus, changing, the light intensity of the *candle lamp* remains stable (until blown out). However, the differences between the two concepts are independent of representation, i.e., there is no significant interaction effect. Nevertheless, the apparent effect of representation on two dimensions of the *Interaction Vocabulary* suggests that assessments on the level of actual operations require first-hand experience.

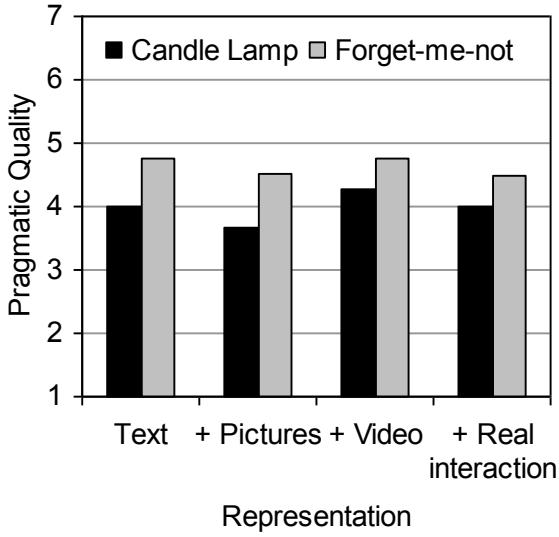

Figure 3: Mean pragmatic quality of the two concepts for each representation

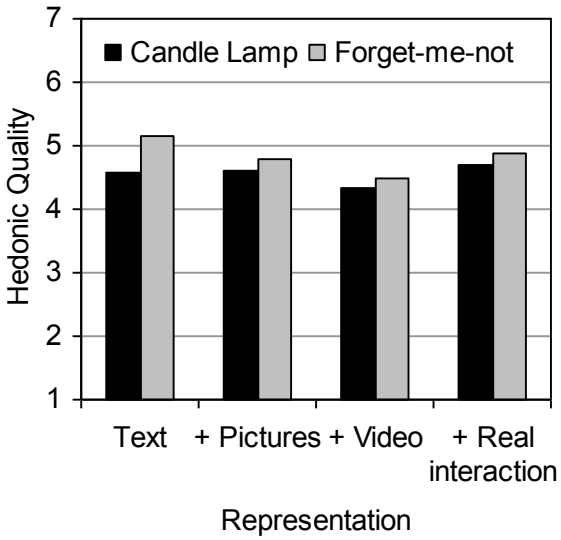

Figure 4: Mean hedonic quality of the two concepts for each representation

CONCLUSION

Neither the global evaluation (i.e., goodness) nor pragmatic and hedonic quality perceptions were significantly influenced by the representation of the respective concept. However, plausible significant differences between the two concepts emerged, which rules out a simple general non-reactivity of the employed measures. The *AttrakDiff2* questionnaire was able to detect differences between concepts but was unaffected by differences in the way concepts were presented. On the level of concrete interaction and its aesthetics, two dimensions of the *Interaction Vocabulary* were affected not only by the concepts but also by representation. As interaction is time-based *per se*, its characteristics may only be disclosed in representations that support the conveyance of according information, which in consequence, leads to assessments depending on the representation. Thus, while general product evaluation and high-level product perception are less susceptible to the way a concept is (re)presented, the assessment on the interaction level requires representation formats that convey the according information (here: time-based attributes of an interaction).

The present study is certainly limited. One specific aspect of the study design, however, may be especially important to explain the null finding concerning representation. We used the textual description as a baseline, which was then enriched by different forms of additional information, such as a video. While this study design definitely reduces the likelihood of finding differences between representations, it appeared to be the most ecologically valid. From our practical work in industry, we learned that a textual description is the most basic and common representation of concepts in early development phases. Future studies will certainly use alternative study designs and additional measures and concepts, in order to develop an exhaustive picture of how a concept's representation impacts various aspects of product evaluation and perception.

We believe that the present research addresses an important but surprisingly neglected issue. So far, academics and practitioners of HCI alike seem to simply assume that a proper empirical product evaluation requires a first-hand experience. Textual representations or mere pictures of products are eyed suspiciously – at least this is a common critical comment on studies that rely on humans' power of imagination, and the ability to develop a rich internal model of a concept, based on only "impoverished" descriptions. The present failure to demonstrate the impact of representation at least on high-level measures is by no means a proof that first-hand experience is unimportant. However, research is needed to differentiate when high fidelity representations are needed and when not, and to qualify, which representation is best for which measure of empirical product evaluation.

ACKNOWLEDGMENTS

This work was supported by the German Federal Ministry of Education and Research (BMBF), project FUN-NI (Grant: 01 IS 09007). For more information, see http://www.fun-ni.org.

REFERENCES

1. Daft, R.L. and Lengel, R. H. Organizational information requirements, media richness and structural design. *Management science, 32,* 5 (1986), 554-571.
2. Diefenbach, S., Hassenzahl, M., Kloeckner, K., Nass, C. and Maier, A. (2010). Ein Interaktionsvokabular: Dimensionen zur Beschreibung der Ästhetik von Interaktion. In H. Brau, S. Diefenbach, K. Göring, M. Peissner, and K. Petrovic (Eds.) Usability Professionals 2010, 27-32. Stuttgart: Fraunhofer..
3. Hassenzahl, M. The Interplay of Beauty, Goodness, and Usability in Interactive Products. *Human-Computer Interaction*, *19* (2004), 319-349.
4. Ince, I. F., Salman, Y. B., and Yildirim, M. E. A user study: the effects of mobile phone prototypes and task complexities on usability. In Proc. *ICIS 2009*, ACM Press (2009), 300-302.
5. Lim, Y., Pangam, A., Periyasami, S., and Aneja, S. Comparative analysis of high- and low-fidelity prototypes for more valid usability evaluations of mobile devices. In *Proc. NordiCHI 2006*, ACM Press (2006), 76-85.
6. Rudd, J., Stern, K., and Isensee, S. Low vs. high-fidelity prototyping debate. *interactions 3,* 1 (1996), 76-85.
7. Sellen, K. M., Massimi, M. A., Lottridge, D. M., Truong, K. N., and Bittle, S. A. The people-prototype problem: understanding the interaction between prototype format and user group. In *Proc. CHI 2009*, ACM Press (2009), 635-638.